

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 12:55:11 ; Search time 36 Seconds  
(without alignments)  
762.490 Million cell updates/sec

Title: US-09-646-778A-288  
Perfect score: 1063  
Sequence: 1 RUSCATLGSQPHPSRRLT.....DTALLDNMKALKLKLTEL 206

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
A_Geneseq.101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*	175	23
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*	175	23
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*	175	23
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*	175	23
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*	175	23
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*	175	23
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*	175	23
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*	175	23
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*	175	23
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*	175	23
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*	175	23
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*	175	23
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*	175	23
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*	175	23
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*	175	23
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*	175	23
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*	175	23
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*	175	23
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*	175	23
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*	175	23
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*	175	23
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*	175	23
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*	175	23

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1063	100.0	206	AAV73837	Human prostate tum
2	1002	94.3	194	AAV76590	Human ovarian tumo
3	1002	94.3	202	AAV75621	Human colon cancer
4	899	84.6	175	AAW77365	zsig10 polypeptide
5	899	84.6	175	AAW37844	Human XAG growth f
6	899	84.6	175	AAW37872	Human protein comp
7	899	84.6	175	AAW59675	Secreted protein 1
8	899	84.6	175	AAW24070	Human PRO1030 prot
9	899	84.6	175	AAU08804	Breast cancer cell
10	899	84.6	175	AAW72203	Human huXAG-1/CCSG

11	899	84.6	175	23	ABB78994	Human XAG-2 protei
12	899	84.6	175	23	ABG34043	Human pro peptide
13	899	84.6	175	23	AAU74939	Human clone DNA594
14	830	78.1	175	23	AAU83628	Human PRO protein,
15	656	61.7	132	21	AAU00109	Human secreted pro
16	605	56.9	131	20	AAV59718	Secreted protein 7
17	598	56.3	116	20	AAV12312	Human 5' EST relat
18	593	55.8	115	21	AAV64672	Human XAG growth f
19	555	52.2	166	19	AAW37846	Breast cancer prot
20	555	52.2	166	21	AAU00194	Breast cancer-asso
21	555	52.2	166	22	AAU25728	Human huXAG-3/CCSG
22	555	52.2	166	22	AAU72205	Amino acid sequenc
23	555	52.2	166	22	AAU31192	Human BCMP 11 poly
24	555	52.2	166	22	AAU07647	BCMP-11 protein.
25	555	52.2	166	23	AAU23228	Human angiogenesis
26	555	52.2	166	23	ABB95520	Human PRO1272 prot
27	555	52.2	166	23	ABB84914	Human PRO protein,
28	555	52.2	166	23	AAU83649	Colon tumour relat
29	555	52.2	166	22	AAU24502	Human ovarian anti
30	555	52.2	184	23	ABP41417	Human ovarian tumo
31	388	36.5	89	20	AAV76568	Human colon cancer
32	361	34.0	180	22	AAU73737	Human 5' EST secre
33	285	26.8	56	20	AAV11881	Human 5' EST secre
34	222	20.9	56	20	AAV11939	Human XAG growth f
35	222	20.9	172	19	AAW37845	Clone HP01766 of a
36	222	20.9	172	20	AAV03236	Clone HP01766 of a
37	222	20.9	172	20	AAV03237	Amino acid sequenc
38	222	20.9	172	20	AAV03230	Xenopus sp embryo
39	222	20.9	172	21	AAU10282	Human PRO1376 prot
40	222	20.9	172	21	AAU24432	Human membrane or
41	222	20.9	172	22	AAW88366	Human huXAG-2/CCSG
42	222	20.9	172	22	AAU72204	Human huXAG-2/CCSG
43	222	20.9	172	23	ABB95536	Human angiogenesis
44	218.5	20.6	172	19	AAW80812	Amino acid sequenc
45	214	20.1	172	23	ABB84930	Human PRO1376 prot

ALIGNMENTS

RESULT 1

AAV73837	AAV73837	standard; Protein: 206 AA.
ID	AAV73837	standard; Protein: 206 AA.
AC	AAV73837	
XX		
XX		
XX		
DT	14-MAR-2000	(first entry)
XX		
DE	Human prostate tumor EST fragment derived protein #24.	
XX		
XX	Pancreas; tumor; EST; expressed sequence tag; human; cytostatic; treatment.	
KW		
KW		
OS	Homo sapiens.	
PN	DE19820190-AL.	
XX		
PD	04-NOV-1999.	
XX		
PF	28-APR-1998; 98DE-1020190.	
XX		
PR	28-APR-1998; 98DE-1020190.	
XX		
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.	
XX		
PI	Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;	
XX		
DR	WPI; 1999-621386/54.	
DR	N-PSDB; AAZ52865.	
XX		
PT	New human nucleic acid sequences from pancreatic tumors, and related proteins	
XX		



\_\_\_\_\_

FT Peptide 138...150  
 FT /label= Epitope  
 FT /note= "Claim 12"  
 PN WO9807749-A1.  
 XX 26-FEB-1998.  
 XX 22-AUG-1997; 97WO-US14139.  
 XX 23-AUG-1996; 96WO-US13766.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Dillion PJ, Ebner R, Endress GA, Yu G;  
 XX WPI: 1998-169093/15.  
 XX N-PSDB; AAV19155.  
 XX New isolated human XAG growth factor(s) - used to develop products  
 XX for treating e.g. liver, lung or breast diseases or  
 XX hyperproliferative disorders, e.g. cancer.  
 XX Claim 1; Fig 1; 141pp; English.  
 XX This polypeptide comprises huxAG-1, a member of a novel family of  
 CC human growth factors also including huxAG-2 (see AAW37845) and  
 CC huxAG-3 (see AAW37846). These proteins share homology with the  
 CC XAG protein of *Xenopus laevis*, which is involved in embryogenesis  
 CC and is expressed in adult tissue. huxAG-1 is specifically found  
 CC in cancerous colon cells and may therefore be a growth factor for  
 CC colon cancer. huxAG-1 cDNA (see AAV19155) was isolated from a cDNA  
 CC library derived from human colon cancer tissue. Vectors, host  
 CC cells, antibodies, and screening methods for identifying agonists  
 CC and antagonists of huxAG-1 are provided. huxAG polypeptides are  
 CC growth factors and can be used to stimulate proliferation of cells.  
 CC They can be used to stimulate the proliferation and differentiation  
 CC of hepatocytes to alleviate or treat liver diseases and pathologies  
 CC such as fulminant liver failure caused by cirrhosis, liver damage  
 CC caused by viral hepatitis and toxic substances. They can also be  
 CC used to stimulate or promote liver regeneration, e.g. after  
 CC surgery. They can also be used to prevent and heal damage to the  
 CC lungs caused by various pathological states. They can be used to  
 CC stimulate proliferation and differentiation and promote the repair  
 CC of alveoli and broncholar epithelium to prevent, attenuate, or  
 CC treat acute or chronic lung damage, e.g. emphysema, which results  
 CC in the progressive loss of alveoli, and inhalation injuries, e.g.  
 CC resulting from smoke inhalation and burns, that cause necrosis of  
 CC the broncholar epithelium and alveoli. They can also be used to  
 CC stimulate the proliferation and differentiation of breast tissue  
 CC and could therefore be used to promote healing of breast tissue  
 CC injury due to surgery, trauma or cancer. Antagonists can be used  
 CC to treat hyperproliferative disorders, including cancer, in  
 CC particular hepatocellular carcinoma, osteoclastoma, breast cancer,  
 CC or colon cancer. The products can also be used for detection and  
 CC diagnosis.

Query Match 84.6%; Score 899; DB 19; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 MEKIPVSFALLLVSYTLARDTTVPKAKDKTDSRPKLPQTLSRGWGDLIWTQYEE 91  
 DB 1 MEKIPVSFALLLVSYTLARDTTVPKAKDKTDSRPKLPQTLSRGWGDLIWTQYEE 60

OY 92 ALYKSTSNKPLMIHHLDECPSHQALKKVFENKETQKLAEOFVLLNLVYETTDKHLSP 151  
 DB 61 ALYKSTSNKPLMIHHLDECPSHQALKKVFENKETQKLAEOFVLLNLVYETTDKHLSP 120

OY 152 DQGYVPRIMFVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 206  
 DB 152 DQGYVPRIMFVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 175

RESULT 7  
 AAY59675

Db 121 DQGYVPRIMFVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 175

RESULT 6  
 AAW37872  
 ID AAW37872 standard; Protein: 175 AA.  
 XX  
 AC AAW37872;  
 XX  
 DT 10-AUG-1998 (first entry)  
 XX  
 DE Human protein comprising secretory signal amino acid sequence 9.  
 XX  
 KW Human protein; secretory signal; nutritional source; cytokine;  
 KW immunity/haematopoiesis; activin; inhibin; tumour; chemotactic;  
 KW chemokinetic; thrombolytic; anti-inflammatory; inhibition;  
 KW stomach cancer cell.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 PN WO9811217-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 XX 12-SEP-1997; 97WO-JP03239.  
 PF  
 XX 13-SEP-1996; 96JP-0243060.  
 PR  
 XX (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENTRE.  
 XX  
 XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 DR WPI: 1998-207380/18.  
 XX N-PSDB; AAV29047, AAV29048.  
 XX Human proteins with secretory signal sequences - used to treat  
 PT immune deficiencies, infections, tumours, and haematopoietic  
 PT disorders, etc.  
 XX  
 XX Claim 1; Pages 79; 131pp; English.  
 XX This is the amino acid sequence of a novel human protein comprising  
 CC a secretory signal isolated from stomach cancer cells. Its proteins  
 CC can be used as nutritional sources or supplements. The proteins may  
 CC also have cytokine functions, immune modulating functions,  
 CC haematopoiesis regulating activity, activin/inhibin regulating  
 CC activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity.

Query Match 84.6%; Score 899; DB 19; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 MEKIPVSFALLLVSYTLARDTTVPKAKDKTDSRPKLPQTLSRGWGDLIWTQYEE 91  
 DB 1 MEKIPVSFALLLVSYTLARDTTVPKAKDKTDSRPKLPQTLSRGWGDLIWTQYEE 60

OY 92 ALYKSTSNKPLMIHHLDECPSHQALKKVFENKETQKLAEOFVLLNLVYETTDKHLSP 151  
 DB 61 ALYKSTSNKPLMIHHLDECPSHQALKKVFENKETQKLAEOFVLLNLVYETTDKHLSP 120

OY 152 DQGYVPRIMFVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 206  
 DB 152 DQGYVPRIMFVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 175

RESULT 7  
 AAY59675

ID AAY59675 standard; Protein: 175 AA.  
 AC AAY59675;  
 XX  
 DT 18-JAN-2000 (first entry)  
 XX  
 DE Secreted protein 108-008-5-0-A6-FL.  
 XX  
 KW Secreted protein; fingerprint identification technique;  
 KW chromosome mapping; human; hereditary disease; diagnosis; cancer;  
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;  
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;  
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;  
 KW hypertension.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9940189-A2.  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 09-FEB-1999; 99WO-IB00282.  
 XX  
 PR 09-FEB-1998; 98US-0074121.  
 PR 13-APR-1998; 98US-0081563.  
 PR 10-AUG-1998; 98US-0096116.  
 PR 04-SEP-1998; 98US-0099273.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 XX Bougueleret L, Duclert A, Dumas Milne Edwards J;  
 PI  
 DR WPI; 1999-600966/51.  
 DR N-PSDB; AA240803.  
 XX  
 XX Extended cDNAs useful for expressing secreted proteins and to obtain  
 PT specific antibodies -  
 XX  
 PS Claim 10; Page 199; 244pp; English.  
 XX  
 CC This sequence represents a human secreted protein of the invention.  
 CC The extended cDNAs (or genomic DNAs obtainable from them) may be used to  
 CC prepare PCR primers and probes. These are useful for forensic matching or  
 CC positive identification by DNA sequencing. They may also be used in  
 CC alternative fingerprint identification techniques. Antibodies against the  
 CC proteins encoded by the extended cDNAs are useful in identification of  
 CC tissue types or cell species, as well as identifying tissue specific  
 CC soluble proteins. The sequences can be used for chromosome mapping and  
 CC identification of genes associated with hereditary diseases or drug  
 CC response. Signal sequences from the cDNAs can be used in construction of  
 CC secretion vectors. Other sequences derived from the extended cDNAs can be  
 CC used to clone upstream genomic DNA sequences including promoters. This is  
 CC in turn useful for identifying proteins that interact with promoter  
 CC sequences. Some of the proteins may be useful in diagnosing and treating  
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,  
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and  
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,  
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.  
 XX  
 SQ Sequence 175 AA;  
 Best Match 84.6%; Score 899; DB 20; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Query 32 MEKIPVSALLLVALSYTLARDTTKPGAKDKTKSRPKLPQTLRSRGWDQLWTQTYEE 91  
 Db 1 MEKIPVSALLLVALSYTLARDTTKPGAKDKTKSRPKLPQTLRSRGWDQLWTQTYEE 60  
 Query 92 ALYKSKTSNKNPLMITHLDECPSQALKKVFRAENKEIQKLAEOFVLLNLVYETDKHLSP 151  
 Db 61 ALYKSKTSNKNPLMITHLDECPSQALKKVFRAENKEIQKLAEOFVLLNLVYETDKHLSP 120

QY 152 DGQYVPRIMFVDPSTVRADITGRYSNRLIYAYEPADTALLDNMKALKLLKTEL 206  
 Db 121 DGQYVPRIMFVDPSTVRADITGRYSNRLIYAYEPADTALLDNMKALKLLKTEL 175  
 RESULT 8  
 AAB24070  
 ID AAB24070 standard; Protein: 175 AA.  
 XX  
 AC AAB24070;  
 XX  
 DT 29-JAN-2001 (first entry)  
 XX  
 DE Human PRO1030 protein sequence SEQ ID NO:40.  
 XX  
 KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;  
 KW proliferation; tumorigenesis; identification; cancer; cytostatic;  
 KW neotropic; neuroprotective; antiinflammatory; immunosuppressive;  
 KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;  
 KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;  
 KW hypothalamic disorder; glandular disorder; macrophagal disorder;  
 KW epithelial disorder; stromal disorder; blastocoealic disorder;  
 KW inflammatory disorder; immunologic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200053755-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00376.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 07-JUL-1999; 99US-0143048.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 20-DEC-1999; 99WO-US30911.  
 PR 05-JAN-2000; 2000WO-US00219.  
 XX  
 XX (GETH ) GENENTECH INC.  
 XX  
 XX Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;  
 PI Watanabe CK, Wood WI;  
 XX  
 XX WPI; 2000-572270/53.  
 DR N-PSDB; AAC58380.  
 XX  
 PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the  
 PT treatment, diagnosis and prevention of cancer -  
 XX  
 PS Claim 61; Fig 28; 286pp; English.  
 XX  
 CC The present invention describes an isolated antibody that binds to  
 CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,  
 CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,  
 CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,  
 CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,  
 CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell  
 CC growth. The PRO polypeptides and nucleotides are useful in the  
 CC treatment, diagnosis and prevention of cancer. The antibodies and other  
 CC anti-tumour compounds may be used to treat various conditions, including  
 CC those characterised by overexpression and/or activation of the amplified  
 CC PRO genes. Exemplary conditions or disorders to be treated with such  
 CC antibodies and other compounds include benign or malignant tumours  
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,  
 CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic  
 CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),  
 CC leukaemias and lymphoid malignancies, other disorders such as neuronal,  
 CC glial, astrocytal, hypothalamic and other glandular, macrophagal,  
 CC epithelial, stromal and blastocoealic disorders, and inflammatory,  
 CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR

CC primers and hybridisation probes used in the isolation of the human PRO  
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human  
CC PRO polynucleotide and protein sequences given in the exemplification of  
CC the present invention.

```

SQ      Sequence      175 AA;

Query Match      84.6%; Score 899; DB 21; Length 175;
Best Local Similarity 100.0%; Pred. No. 2.2e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      32 MEKIPVSFLLLVSLTLLARDITVPGAKKDKTKDSRPKLPQTLSRGWGQOLIWTQYEE 91
      |||||
Ddb      1 MEKIPVSFLLLVSLTLLARDITVPGAKKDKTKDSRPKLPQTLSRGWGQOLIWTQYEE 60
      |||||

QY      92 ALYKSTSNKPLMIIHHLDCPSHQALKKVFAENKETQKLAEQFVLLNLYVETTDKHLSP 151
      |||||
Ddb      61 ALYKSTSNKPLMIIHHLDCPSHQALKKVFAENKETQKLAEQFVLLNLYVETTDKHLSP 120
      |||||

QY      152 DGOYVPRIMFVDSLIVRADITGRYSNRLYAYEPADTALLDDNNMKALKLKTTEL 206
      |||||
Ddb      121 DGOYVPRIMFVDSLIVRADITGRYSNRLYAYEPADTALLDDNNMKALKLKTTEL 175
      |||||

```

RESULT 9	
AAU08804	
ID	AAU08804 standard; Protein; 175 AA.
XX	
AC	AAU08804;
XX	
DT	19-DEC-2001 (first entry)
DE	
XX	Breast cancer cell membrane associated protein 7 (BCMP 7).
XX	
KW	Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
XW	metastasis; cytostatic; gene therapy; antibody; antisense;
XW	chromosome 7p21.3; vaccine.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
Peptide	1..20
TT	/note= "Signal peptide"
Protein	21..175
TT	/note= "Mature BCMP 7"
XX	
XX	WO200163290-A1.
XX	
XX	30-AUG-2001.
XX	
XX	21-FEB-2001; 2001WO-GB00734.
XX	
XX	25-FEB-2000; 2000GB-0004576.
XX	
XX	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX	
XX	Boyd RS, Stamps AC, Terrett JA, Tyson KL;
XX	
XX	WPI; 2001-570651/64.
XX	N-PSDB; AAS13480.
XX	
XX	Diagnosing, preventing and treating breast cancer using a breast cancer
XX	cell membrane protein BCMP 7 -
XX	Claim 1; Fig 1; 62pp; English.

The invention describes the novel use of a protein found in breast cancer cell membranes (BCMP 7) for diagnosing, preventing and treating breast cancers. The peptide has cytostatic action and potential uses in gene therapy and in vaccines. The polypeptide, antisense nucleic acids, or fusion proteins comprising and Green Fluorescent protein or the DsRed Fluorescent-protein, antibodies specific for and/or nucleic acid are used for the prevention and/or treatment of breast cancer. Antibodies

CC	against may also be used for screening for and/or diagnosis of breast cancer in a patient. The method for monitoring/assessing breast cancer treatment in a patient and for the identification of metastatic breast cancer cells in samples from a patient. This sequence is breast cancer cell associated protein 7 (BCMP 7), encoded by a gene located on chromosome 7p21.3, described in the method of the invention.
XX	
SQ	Sequence 175 AA;
	Query Match 84.6%; Score 899; DB 22; Length 175; Best Local Similarity 100.0%; Pred. No. 2.2e-90; Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	32 MEKTPVSAFLLLVSLTLYARDTTVKCAKDDTKDSRPKLPQTLSRGWGDLIWTQYEE 91           Db 1 MEKTPVSAFLLLVSLTLYARDTTVKCAKDDTKDSRPKLPQTLSRGWGDLIWTQYEE 60 
QY	92 ALYKSSTSNRPLMIIHHLDECPSOALKVFAENKETQLAEQFVLNLNLYETTDKLSP 151           Db 61 ALYKSSTSNRPLMIIHHLDECPSOALKVFAENKETQLAEQFVLNLNLYETTDKLSP 120 
QY	152 DGQYVPRIMFDVSLTVRADITGRYSNRLYAYPADTALLDNNMKALKLKTLE 206 
Dd	121 DGQYVPRIMFDVSLTVRADITGRYSNRLYAYPADTALLDNNMKALKLKTLE 175 

RESULT 10	
AAB72203	
ID	AAB72203 standard; Protein; 175 AA.
XX	
XX	AAB72203;
DT	04-MAY-2001 (first entry)
XX	
DE	Human huxAG-1/CCSG colon cancer specific gene amino acid sequence.
XX	
KW	Human; growth factor; huxAG-1; colon cancer specific gene; CCSG;
KW	cell proliferation; liver disease; fulminant liver failure; cirrhosis;
KW	hepatitis; cancer; colon cancer; colorectal carcinoma.
XX	
OS	Homo sapiens.
XX	
PN	US6171816-B1.
XX	
PD	09-JAN-2001.
XX	
PF	22-AUG-1997; 97US-0916576.
XX	
PR	23-AUG-1996; 96US-0024347.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Yu G, Dillon PJ, Ebner R, Endress GA;
XX	
DR	WPI; 2001-136872/14.
XX	
DR	N-PSDB; AAF63314.
XX	
PPT	Novel human growth factor polypeptide useful for diagnosing and
PPT	treating colon cancer and liver diseases, to prevent and heal damage to
PPT	the lungs and for identifying modulators of therapeutic use -
XX	
PS	Claim 165; Fig 1; 59pp; English.

This invention relates to a human growth factor polypeptide huXAG-1 also known as a colon cancer specific gene (CCSG). HuXAG-1 stimulates cell proliferation as a growth factor. The HuXAG-1 protein is useful for identifying compounds capable of enhancing or inhibiting cellular response induced by huXAG-1. The protein is also useful for stimulating proliferation of cells e.g. colon, breast, liver and lung cells, and hepatocytes. It is useful for alleviating or treating liver diseases and pathologies such as fulminant liver failure caused by cirrhosis, liver damage caused by viral hepatitis and toxic substances, for preventing and treating damage to the lungs caused by various pathological states and

CC for promoting healing of breast tissue injury due to surgery, trauma or  
 CC cancer. HuXAG-1 and the identified antagonist are useful for treating  
 CC cancer, in particular colon cancer. Detecting altered levels of huXAG-1  
 CC and its polynucleotides are useful for diagnosing or detecting cancer in  
 CC mammals. The gene encoding huXAG-1 is useful for monitoring human  
 CC colorectal carcinoma. huXAG-1 nucleic acid molecules are also useful for  
 CC chromosome identification. The present sequence represents the huXAG-1  
 CC protein.  
 CC  
 XX  
 SQ

Query Match 84.6%; Score 899; DB 22; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 MEKIPVSFALLLVALLSYTLARDTTVPKAKKDTKDSRPKLPQTLRSRGWDQLIWTQYEE 91  
 DB 1 MEKIPVSFALLLVALLSYTLARDTTVPKAKKDTKDSRPKLPQTLRSRGWDQLIWTQYEE 60  
 QY 92 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIOKLAEOQVLLNLVYETTDKHLSP 151  
 DB 61 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIOKLAEOQVLLNLVYETTDKHLSP 120  
 QY 152 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDDNMKKALKLKTTEL 206  
 DB 121 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDDNMKKALKLKTTEL 175

RESULT 11  
 ABB78994  
 ID ABB78994 standard; Protein; 175 AA.

AC ABB78994;

DT 02-AUG-2002 (first entry)

XX Human XAG-2 protein sequence SEQ ID NO:4473.

XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
 KW genetic analysis; diagnostic; antisense therapy.

XX Homo sapiens.

XX WO200229086-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30732.

XX 02-OCT-2000; 2000US-237271P.

XX (FARB ) BAYER CORP.

XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
 PI Thialingam A, Lewis ME;

XX WPI; 2002-426115/45.

XX N-PSDB; ABQ60777.

XX New isolated nucleic acid that is differentially expressed in cancer  
 PT tissues useful for determining the presence of colon cancer in a cell  
 PT or tissue type, and in antisense therapy

XX Claim 5; Fig 3; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which

CC hybridises to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists.  
 CC  
 XX  
 SQ

Sequence 175 AA;

Query Match 84.6%; Score 899; DB 23; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 MEKIPVSFALLLVALLSYTLARDTTVPKAKKDTKDSRPKLPQTLRSRGWDQLIWTQYEE 91  
 DB 1 MEKIPVSFALLLVALLSYTLARDTTVPKAKKDTKDSRPKLPQTLRSRGWDQLIWTQYEE 60  
 QY 92 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIOKLAEOQVLLNLVYETTDKHLSP 151  
 DB 61 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIOKLAEOQVLLNLVYETTDKHLSP 120  
 QY 152 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDDNMKKALKLKTTEL 206  
 DB 121 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDDNMKKALKLKTTEL 175

RESULT 12  
 ABG34043

ID ABG34043 standard; Protein; 175 AA.

XX AC ABG34043;

XX 15-JUL-2002 (first entry)

XX Human Pro peptide #14.

XX Human; PRO: secreted protein; transmembrane protein;  
 KW genetic disorder; tumour; cancer.

XX Homo sapiens.

XX WO200224888-A2.

XX 28-MAR-2002.

XX 29-AUG-2001; 2001WO-US27099.

XX 01-SEP-2000; 2000US-229896P.

XX 05-SEP-2000; 2000US-230621P.

XX 22-SEP-2000; 2000US-235147P.

XX 10-NOV-2000; 2000WO-US30873.

XX 12-JAN-2001; 2001US-261878P.

XX 16-JAN-2001; 2001US-261910P.

XX 16-JAN-2001; 2001US-261939P.

XX 25-JAN-2001; 2001US-262150P.

XX 02-FEB-2001; 2001US-266421P.

XX 03-FEB-2001; 2001US-267623P.

XX 28-FEB-2001; 2001WO-US06520.

XX 09-MAR-2001; 2001US-274399P.

XX 03-APR-2001; 2001US-280982P.

XX 04-APR-2001; 2001US-282129P.

XX 09-MAY-2001; 2001US-290589P.

XX 25-MAY-2001; 2001WO-US17092.

XX 01-JUN-2001; 2001WO-US17800.

XX 20-JUN-2001; 2001WO-US19692.

XX 29-JUN-2001; 2001WO-US21066.

XX 09-JUL-2001; 2001WO-US21735.

XX (GETH ) GENENTECH INC.

XX Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC, Gurney AL, Smith V, Watanabe CK, Wood WI, Zhang Z; Fong S;  
XX  
XX  
XX  
XX WPI: 2002-362426/39.  
XX N-PSDB; ABK69974.  
XX  
XX New PRO polypeptides and polynucleotides encoding the polypeptides, useful in gene therapy, chromosome identification, tissue typing, or for genetic analysis of individuals with genetic disorders -  
XX  
XX  
XX  
XX Claim 11; Figure 28; 218pp; English.  
XX  
XX This invention relates to the cDNA and protein sequences of novel secreted and transmembrane polypeptides PRO polypeptides. The invention also comprises a method for producing the proteins of the invention by recombinant means and antibodies specific for the protein of the invention. The antibody may be used for detecting the PRO proteins of the invention and may be used to modify their activity.  
XX polynucleotides may be used as hybridisation probes for a cDNA library to isolate the full-length PRO cDNA or to isolate other cDNAs, to construct hybridisation probes for mapping the gene which encodes that PRO and for genetic analysis of individuals with genetic disorders, in assays to identify other proteins or molecules involved in binding reaction, to generate transgenic animals or knock-out animals which turn are useful in the development and screening of therapeutically useful reagents, for chromosome identification, and tissue typing. The PRO polypeptides are useful in gene therapy, and as molecular weight markers for protein electrophoresis purposes. The sequences may also be used to detect overexpression on PRO polypeptides in cancerous tumours and for screening for differentially expressed genes using microarray technology. The present sequence represents a human PRO protein of the invention.  
XX  
XX  
XX Sequence 175 AA;

Query Match 84.6%; Score 899; DB 23; Length 175;  
Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 32 MEKIPVSFLLVALSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLWTQTYEE 91  
Db 1 MEKIPVSFLLVALSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLWTQTYEE 60  
QY 92 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIQKLAEQVLLNLVYETTDKHLSP 151  
Db 61 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIQKLAEQVLLNLVYETTDKHLSP 120  
QY 152 DGQYVPRIMFVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 206  
Db 121 DGQYVPRIMFVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 175

RESULT 13  
AAU74939  
ID AAU74939 standard; Protein; 175 AA.  
XX  
XX  
XX  
XX 23-APR-2002 (first entry)  
XX  
XX Human clone DNA59485-1336 amino acid sequence of PRO1030 protein.  
XX  
XX Human; clone DNA59482-1336; PRO1030; immune response modulator;  
KW PRO polypeptide; immune related disorder; systemic lupus erythematosus;  
KW rheumatoid arthritis; osteoarthritis; systemic sclerosis; sarcoidosis;  
KW autoimmune thrombocytopenia; thyroiditis; diabetes mellitus; anaemia;  
KW multiple sclerosis; inflammatory bowel disease; allergic rhinitis.  
OS Homo sapiens.  
XX  
XX Key .Location/Qualifiers

FT Peptide 1..20  
FT /label= Signal\_peptide  
FT Protein 21..175  
FT /label= Mature\_PRO1030\_protein  
FT Modified-site 30..34  
FT /label= Phosphorylation\_site  
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"  
XX  
XX  
XX  
XX WO200192331-A2.  
XX  
XX 06-DEC-2001.  
XX  
XX 25-MAY-2001; 2001WO-US17092.  
XX  
XX 30-MAY-2000; 2000WO-US14941.  
XX 02-JUN-2000; 2000WO-US1264.  
XX 01-DEC-2000; 2000WO-US32678.  
XX (GETH ) GENENTECH INC.  
XX  
XX Eaton DL, Fong S, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL; Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI: 2002-075461/10.  
XX N-PSDB; ABK14011.  
XX  
XX Isolated PRO221, PRO617, PRO1030, PRO4302 polypeptides, useful for treating immune disorders such as thyroiditis, diabetes mellitus, allergic disease, asthma, allergic rhinitis, atopic dermatitis -  
XX  
XX Claim 10; Fig 6; 104pp; English.

XX The present invention relates to a new polypeptide having at least 80% amino acid sequence identity to fully defined amino acid sequence of PRO221 (AAU74937), PRO617 (AAU74938), PRO1030 (AAU74939) or PRO4302 (AAU74940) as given in specification. The PRO polypeptides of the invention are useful for treating an immune related disorder such as systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis, systemic sclerosis, sarcoidosis, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus, multiple sclerosis, inflammatory bowel disease, allergic rhinitis and anaemia. The methods of the invention are useful for identifying an antisense nucleic acid and for enhancing the infiltration of inflammatory cells into the tissue of a mammal. The anti-PRO antibodies are useful for determining the presence of PRO polypeptides and for diagnosing an immune-related disease in a mammal. The molecules of the invention are useful for identifying a compound that inhibits the activity of PRO polypeptides or expression of a gene encoding the PRO polypeptide. PRO polypeptides are also useful for increasing or decreasing the infiltration of inflammatory cells from the vasculature into a tissue of a mammal, increasing or decreasing activity or proliferation of T-lymphocytes. The present amino acid sequence represents the PRO1030 polypeptide of the invention. The PRO1030 sequence was isolated from the clone DNA59485-1336.

XX Sequence 175 AA;  
XX  
XX Query Match 84.6%; Score 899; DB 23; Length 175;  
XX Best Local Similarity 100.0%; Pred. No. 2.2e-90;  
XX Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 32 MEKIPVSFLLVALSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLWTQTYEE 91  
Db 1 MEKIPVSFLLVALSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLWTQTYEE 60  
QY 92 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIQKLAEQVLLNLVYETTDKHLSP 151  
Db 61 ALYKSTSNKPLMIHLLDECPHSQALKKVFPAENKEIQKLAEQVLLNLVYETTDKHLSP 120  
QY 152 DGQYVPRIMFVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 206  
Db 121 DGQYVPRIMFVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 175





```

xx
SQ Sequence 132 AA;
Query Match 61.7%; Score 656; DB 21; Length 132;
Best Local Similarity 100.0%; Pred. No. 7.2e-64;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 32 MEKIPVSFLLVLSYTLARDTTVKPGAKKDTKDSRPKLPQTLNRGWDQLIWTQTYEE 91
Db 1 MEKIPVSFLLVLSYTLARDTTVKPGAKKDTKDSRPKLPQTLNRGWDQLIWTQTYEE 60
Qy 92 ALYKSKTSNKPLMIHHLDCEPHSQALKKVFPAENKEIQKLAEOFLNLYVETTDKHLSP 151
Db 61 ALYKSKTSNKPLMIHHLDCEPHSQALKKVFPAENKEIQKLAEOFLNLYVETTDKHLSP 120
Qy 152 DGQYVP 157
Db 121 DGQYVP 126

```

Search completed: January 17, 2003, 12:55:56  
Job time : 37 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 12:55:16 ; Search time 20 Seconds  
(without alignments)  
990.185 Million cell updates/sec

Title: US-09-646-778A-288  
Perfect score: 1063  
Sequence: 1 RLSCAGTSLGSGPHSRRLT.....DTALLDNMKALKLLKTEL 206

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	899	84.6	175	2 JE0350	Anterior gradient-
2	199	18.7	186	2 T31643	hypothetical prote
3	119	11.2	150	2 D69100	probable protein d
4	87.5	8.2	458	2 S50919	hypothetical prote
5	87	8.2	348	2 G72016	probable disulfide
6	87	8.2	348	2 C86607	disulfide bond iso
7	85.5	8.0	1830	1 S19188	myosin-v - chicken
8	84.5	7.9	1151	2 A61717	probable peptidogl
9	84	7.9	771	2 S51421	hypothetical prote
10	83.5	7.9	217	2 T37859	probable transcrip
11	83.5	7.9	918	2 C96829	unknown protein F1
12	83	7.8	479	2 T40683	cell cycle protein
13	83	7.8	507	2 S56143	cell cycle protein
14	82.5	7.8	788	2 C70441	lipocate-protein li
15	81.5	7.7	893	2 G59431	phosphatidylinosit
16	81.5	7.7	970	2 S29069	inositol polyphosp
17	81.5	7.7	1158	2 T50454	probable rho1 GDP-
18	81.5	7.7	1397	2 T10466	DNA topoisomerase
19	81	7.6	610	2 T22909	hypothetical prote
20	81	7.6	635	2 T23874	hypothetical prote
21	80	7.5	154	2 F82633	probable protein d
22	80	7.5	259	2 A96740	F14023.14 (importe
23	80	7.5	843	2 S53947	hypothetical prote
24	79.5	7.5	277	2 T37629	hypothetical protei
25	79.5	7.5	301	2 G71929	GTP-binding protei
26	79.5	7.5	642	2 F83718	ABC transporter (A
27	79.5	7.5	921	2 F84593	hypothetical prote
28	79	7.4	477	2 T37791	probable rna polym
29	79	7.4	510	1 S43516	carboxypeptidase C

30	78.5	7.4	280	2 A69590	shikimate 5-dehydr
31	78.5	7.4	917	2 T05430	hypothetical prote
32	78.5	7.4	1033	2 A96714	hypothetical prote
33	78.5	7.4	1262	2 I48855	tripeptidyl-peptid
34	78.5	7.4	1570	2 AC2012	hypothetical prote
35	78	7.3	337	2 S06956	segmentation prote
36	78	7.3	1828	2 B59254	myosin heavy chain
37	78	7.3	1853	1 A46761	myosin heavy chain
38	78	7.3	1855	2 A59254	formate dehydrogen
39	76.5	7.2	985	2 E69850	RNA-directed DNA p
40	76.5	7.2	1025	2 T10259	thioredoxin disulf
41	76	7.1	166	2 D72018	thioredoxin disulf
42	76	7.1	166	2 D86606	probable uracil ki
43	76	7.1	367	2 T39574	oligopeptide-bindi
44	76	7.1	543	1 F64871	probable peptide A
45	76	7.1	583	2 E75529	

ALIGNMENTS

RESULT 1

JE0350

Anterior gradient-2 - human

C:Species: Homo sapiens (man)

C:Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 21-Jul-2000

C:Accession: JE0350

R:Thompson, D.A.; Weigel, R.J.

Biochem. Biophys. Res. Commun. 251, 111-116, 1998

A:Title: hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is

A:Reference number: JE0350; MUID:99009231; PMID:9790916

A:Accession: JE0350

A:Molecule type: mRNA

A:Residues: 1-175 <THO>

A:Cross-references: GB:AF007791; NID:g3779196; PIDN:AAC77358.1; PID:g3779197

C:Comment: This protein is coexpressed with estrogen receptor (ER) in breast cancer c

C:Genetics:

A:Gene: hAG-2

C:Superfamily: human anterior gradient-2

Query Match 84.6%; Score 899; DB 2; Length 175;

Best Local Similarity 100.0%; Pred. No. 3.1e-74;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 MEKIPVSAPLLVALSYTLARDTVKPGAKKDKDSRPKLPQTLSRGWDQLIWTQYEE 91

Db 1 MEKIPVSAPLLVALSYTLARDTVKPGAKKDKDSRPKLPQTLSRGWDQLIWTQYEE 60

Qy 92 ALYKSKTSNKPLMIHHLDPCPHSQALKKVFENKEIQKLAEQFVLLMLVYETTDKHLSP 151

Db 61 ALYKSKTSNKPLMIHHLDPCPHSQALKKVFENKEIQKLAEQFVLLMLVYETTDKHLSP 120

Qy 152 DQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKALKLLKTEL 206

Db 121 DQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKALKLLKTEL 175

RESULT 2

T31643

hypothetical protein Y57A10A.u - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T31643

R:Smyle, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21048

A:Accession: T31643

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-186 <WIL>

A:Cross-references: EMBL:AL117195; PIDN:CAB55026.1; CESP:Y57A10A.u

A:Experimental source: clone Y57A10A

C:Genetics:

A:Gene: CESP:Y57A10A.u  
A:Introns: 70/1  
C:Superfamily: human anterior gradient-2

Query Match 18.7%; Score 199; DB 2; Length 186;  
Best Local Similarity 30.0%; Pred. No. 1.1e-10;  
Matches 51; Conservative 38; Mismatches 57; Indels 24; Gaps 6;

QY 41 LLLVALSYTLARDTTVPKAKKDTKDSRPKLPQTLRSRGWGLQTLQTYEEALYKSKTSN 100  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 4 LLLVAL-----VSASAYASFDKIKDSIQNPLRGFGDDIAWK-WEDIAETALDFD 53

QY 101 KPLMIHHLDPCPSQALKKFAEN---KEIQKLAQFVLLNLV--YETDKHLSPDQGY 155  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 54 KPILLIHKSMCHACKALKKTFOQSNNAKKAFFKLSERFVWNTEDDDPFEERYPDGKY 113

QY 156 VPRIMFVDPSTLVRADITGRYSNRL-----YAYEPADTALLLDNNKKALK 200  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 114 IPRULFDKN-----GDLQEPKNKAQYKKNYAYYYSPADILNSMKDVLK 159

RESULT 3  
D69100  
probable protein disulfide-isomerase (EC 5.3.4.1) MTH1745 [similarity] - Methanobacteriu

C:Species: Methanobacterium thermoautotrophicum  
C:Date: 03-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Sep-2000  
C:Accession: D69100

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: D69100  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-150 <MTH>

A:Cross-references: GB:AE000930; GB:AE000666; NID:g2622872; PIDN:AAB86215.1; PID:g262287  
A:Experimental source: strain Delta H

C:Genetics:  
A:Gene: MTH1745  
C:Keywords: intramolecular oxidoreductase; isomerase

Query Match 11.2%; Score 119; DB 2; Length 150;  
Best Local Similarity 27.1%; Pred. No. 0.0015;  
Matches 35; Conservative 28; Mismatches 46; Indels 20; Gaps 4;

QY 39 AFLLLVALSYTLARDTTVPKAKKDTKDSRPKLPQTLRSRGWGLQTLQTYEEALYKSKT 98  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 19 AGLSLVLIITV-----QPRVPSLTDEKO-LKWYTEHDEAIKEASR 60

QY 99 SNKPLMIHHLDPCPSQALKKFAENKEIQ-KLAQFVLLNLVYETDKHLSPDQGY-V 156  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 61 TGNKVFVFSASWCPACQKLESETLQNTQVORLAEFDIAVKIDVTSPLSSRYIYGV 120

QY 157 PRIMFVDPSS 165  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 121 PTVIIDPS 129

RESULT 4  
S50919  
hypothetical protein YMR191w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR191w  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 19-Apr-2002

C:Accession: S50919; S55721  
R:Pearson, D.; Bowman, S.  
submitted to the EMBL Data Library, January 1995

A:Reference number: S50919  
A:Accession: S50919  
A:Molecule type: DNA

A:Residues: 1-158 <PEA>

A:Cross-references: EMBL:Z47815; NID:g642280; PIDN:CAA87812.1; PID:g642283; MIPS:YMR1  
R:Gangloff, S.; McDonald, J.P.; Bendixen, C.; Arthur, L.; Rothstein, R.  
Mol. Cell. Biol. 14, 8391-8398, 1994

A:Title: The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homo  
A:Reference number: A56359; MUID:95059068; PMID:7969174

A:Accession: S55721  
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-127, 'Y', 129-131, 'LAG', 136-249 <GNA>

A:Cross-references: EMBL:U22341; NID:g726277; PIDN:AAB60288.1; PID:g726278  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1995

C:Genetics:  
A:Cross-references: SGD:S0004803  
A:Map position: 13R

C:Superfamily: Saccharomyces cerevisiae hypothetical protein YMR191w

Query Match 8.2%; Score 87.5; DB 2; Length 458;  
Best Local Similarity 20.1%; Pred. No. 4.5;  
Matches 55; Conservative 35; Mismatches 74; Indels 109; Gaps 13;

QY 15 PSRLTQ-----RWVRSRVAMEKIPVSAPLLIVALS----- 47  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 82 PFRNMAVGNNMWMRLMSVHLRQITKSLDRLISLHGNFHSQYNNRNFVWTKSLFE 141

QY 48 -----YTLARDTTVPKAKKDTK-----DSRPKLPOTLSRG-----W---GD 81  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 142 ASTAFRRASCLTVSLTRGRIARDFHPRPVNVSKFASFRVPKCAPRGLFTNNMTTSK 201

QY 82 QLWITQTYEALYK---SKTSNKPLMT-----IHHLDECPHSQ----- 116  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 202 RLIGQRAYSTSIKFTQEAANNNTISLRCFFNSLGLNQCNSCKAYQNASVNTSKQD 261

QY 117 -----ALKVK-----FAENKEIQKLAQFVLLNLVYETDKHLSPDQGYV----- 156  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 262 HVQPVALKKLSQKIDNIRNLELFKMK---TQNEVDETSAIYMEKPGSIEFTISEFNV 319

QY 157 -----PRIMFVDPSTLVRADITGRYSNRLYAYE 184  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |  
Db 320 NGTFSAPLSFLDPSLL--ADLDEMIRN--YKYE 348

RESULT 5  
G72016  
probable disulfide bond isomerase - Chlamydomophila pneumoniae (strains CWL029 and AR39

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C:Accession: G72016; B81523  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: G72016

A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-348 <ARN>

A:Cross-references: GB:AE001673; GB:AE001363; NID:g4377252; PIDN:AAD19071.1; PID:g437  
A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: B81523

A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-348 <REA>

A:Cross-references: GB:AE002251; GB:AE002161; NID:g7189833; PIDN:AAF38711.1; PID:g718  
A:Experimental source: strain AR39, HL cells

C:Genetics:  
A:Gene: Cpn0933; Cpn0928

Query Match 8.2%; Score 87; DB 2; Length 348;  
Best Local Similarity 25.0%; Pred. No. 3.5;

Matches 29; Conservative 26; Mismatches 37; Indels 24; Gaps 4;

Qy 29 RVAMEKIPVSALLLVASYTLARDTTVPKAKDKTDSRPKLPQTLRSRGWGDLIWTQT 88  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 9 RCLSLQKLVATLL--LSLSL--PTLEAAENRDS-----DSIVWHLD 46

Qy 89 YEALYKSTSNKPLMIHHLDEC--PHSQALKKVFPAENKEIOKLAEOQVLLNLVY 142  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 47 YQALQKSKAEALPLLVIFGSDWNGPCMKIRKEVLESPEFIKRVQGVCEVEY 102

RESULT 6  
 C86607  
 disulfide bond isomerase [imported] - Chlamydomophila pneumoniae (strain J138)  
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: C86607  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Ouchi, K.; Shiba, T.; Is  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: C86607  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-348 <STO>  
 A:CROSS-references: GB:BA000008; NID:g8979307; PIDN:BA99141.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: CPj0933

Query Match 8.2%; Score 87; DB 2; Length 348;  
 Best Local Similarity 25.0%; Pred. No. 3.5;  
 Matches 29; Conservative 26; Mismatches 37; Indels 24; Gaps 4;

Qy 29 RVAMEKIPVSALLLVASYTLARDTTVPKAKDKTDSRPKLPQTLRSRGWGDLIWTQT 88  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 9 RCLSLQKLVATLL--LSLSL--PTLEAAENRDS-----DSIVWHLD 46

Qy 89 YEALYKSTSNKPLMIHHLDEC--PHSQALKKVFPAENKEIOKLAEOQVLLNLVY 142  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 47 YQALQKSKAEALPLLVIFGSDWNGPCMKIRKEVLESPEFIKRVQGVCEVEY 102

RESULT 7  
 S19188  
 myosin-v - chicken  
 N:Alternate names: calmodulin-binding protein; myosin I heavy chain, 190K  
 C:Species: Gallus gallus (chicken)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Jan-2001  
 C:Accession: S19188; A44359; B44359; S29249  
 R:Espreafico, E.M.; Cheney, R.E.; Matteoli, M.; Nascimben, A.A.; De-Camilli, P.V.; Lars  
 submitted to the EMBL data library, February 1992  
 A:Description: Complete cDNA coding sequence of chicken brain p190, a calmodulin binding  
 A:Reference number: S19188  
 A:Accession: S19188  
 A:Molecule type: mRNA  
 A:Residues: 1-1830 <ESP>  
 A:CROSS-references: EMBL:Z11718  
 R:Espreafico, E.M.; Cheney, R.E.; Matteoli, M.; Nascimben, A.A.; De-Camilli, P.V.; Lars  
 J. Cell Biol. 119, 1541-1557, 1992  
 A:Title: Primary structure and cellular localization of chicken brain myosin-v (p190), a  
 A:Reference number: A44359; MUID:93107155; PMID:1469047  
 A:Accession: A44359  
 A:Molecule type: mRNA  
 A:Residues: 1-1688, 'R', 1690-1830 <ES2>  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBI:121153, NCBI:121154)  
 A:Note: the codon CGC for residue 1689 is inconsistent with the codon AGC for residue 16  
 A:Accession: B44359  
 A:Molecule type: protein  
 A:Residues: 155-164 <ES3>  
 R:Sanders, G.; Lichte, B.; Meyer, H.E.; Killmann, M.W.  
 FEBS Lett. 311, 295-298, 1992

A:Title: cDNA encoding the chicken ortholog of the mouse dilute gene product. Sequenc  
 A:Reference number: S29249; MUID:93012002; PMID:1383040  
 A:Accession: S29249  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1142, 1144-1830 <SAN>  
 A:CROSS-references: EMBL:X67251; NID:g63364; PIDN:CAA47673.1; PID:g63365  
 C:Comment: The neck domain comprises six approximately 23-residue tandem repeats; thi  
 C:Superfamily: myosin MYO2; myosin motor domain homology  
 C:Keywords: actin binding; ARP; brain; coiled coil; nucleotide binding; P-loop; phosph  
 F:72-752/Domain: myosin motor domain homology <MMOT>  
 F:163-170/Region: actin binding #status predicted  
 F:645-666/Region: neck #status predicted <NEC>  
 F:912-1420/Domain: coiled coil #status predicted <COI>  
 F:1421-1830/Domain: carboxyl-terminal domain #status predicted <CTD>  
 F:169/Binding site: ARP (Lys) #status predicted  
 F:1735/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 8.0%; Score 85.5; DB 1; Length 1830;  
 Best Local Similarity 28.1%; Pred. No. 38;  
 Matches 27; Conservative 21; Mismatches 41; Indels 7; Gaps 3;

Qy 63 DTKDSRPKLPQTLRSRGWGDLIWTQTYEALY-KSTSNKPLMIHHLDECPH--SOALK 119  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 507 DLLDECKMPKGGDDTWAQKLYNTHLNCALFEKPRLSNKAFTIKHFADKVEYQCEGFLE 566

Qy 120 K-----VFAENKEIOKLAEOQVLLNLVYETDKHLSP 151  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 567 KNKDTVYEQIKVLKSSKKFKLLPLFQDEKAISP 602

RESULT 8  
 AG1717  
 probable peptidoglycan bound protein (LPXTG motif) lin2283 [imported] - Listeria inno  
 C:Species: Listeria innocua  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AG1717  
 R:Glaser, P.; Franqueul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
 A:Title: Comparative genomics of Listeria species.  
 A:Reference number: ABI077; MUID:21537279; PMID:11679669  
 A:Accession: AG1717  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1151 <GLA>  
 A:CROSS-references: GB:AL592022; PIDN:CAC97511.1; PID:gl6414795; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Gene: lin2283

Query Match 7.9%; Score 84.5; DB 2; Length 1151;  
 Best Local Similarity 22.8%; Pred. No. 26;  
 Matches 44; Conservative 31; Mismatches 51; Indels 67; Gaps 10;

Qy 50 LARDTTVPKCAK--DTKDSRPKLPQTLRSRGWGDLIWTQ-----TVE 90  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 961 VALDDSVAPYTVQVQDNHKOIASPETLTGAYGEKFTAKQKKIYNTLVKTPANVSGTFN 1020

Qy 91 EA-----LYKSTSNKPLMIHHLDECPHSHQALKKVFPAENKEIOKLAEOQVYL---LNLV 141  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 1021 EKAQTVTFVYQKVTAGN--IIVDIVD-----KNGEKLADSIIVLTGKLNS 1063

Qy 142 YETDKHLSPDGOY-VPRIMFVPSLTVRADITGRYSNR-----LYAYEPADTALL----- 191  
 | : : : : | : | : | : : : : | : : : : | : : : :  
 Db 1064 YRTSAKKISGYKLYQTPK-----NASGFSNTSQRVTYVYEKASTFSISNKG 1111

Qy 192 LDNMKKALKLLKT 204  
 | : : : : | : | : | : : : : | : : : : | : : : :



C:Genetics: SPDB:SPBC776.12c  
A:Gene: SPDB:SPBC776.12c  
A:Map position: 2  
A:Introns: 77/1; 92/2; 331/1  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 7.8%; Score 83; DB 2; Length 479;  
Best Local Similarity 22.4%; Pred. No. 12;  
Matches 28; Conservative 24; Mismatches 55; Indels 18; Gaps 5;

Qy 83 LIWQTYEALYKSTNKP-----LMIHILDECPHSQALKKVFNAEKEIOKLAEOFVLL 138  
Db 352 ILWASCSGASIIYKELRHKPSQERLCDFLEKCLDCKNRISAE-----EALDHDFLYL 407

Qy 139 -NLVETTDKHLSPDQGVPRIMFVDPSTLVTRADITGRYSNRLYAY---EPADTALLLDN 194  
Db 408 DNLAYEKDDDDTAFDNS-----FGETSFEKDEDLTAKHLSHILDFKEQETDEPTSLSK 461

Qy 195 MKKAL 199  
Db 462 RKRSI 466

RESULT 13  
S56143  
C:Species: Schizosaccharomyces pombe  
C:Date: 28-Oct-1995 #sequence\_revision 10-Nov-1995 #text\_change 10-Dec-1999  
C:Accession: S56143; T40682  
R:Masai, H.; Miyake, T.; Arai, K.  
EMBO J. 14, 3094-3104, 1995  
A:Title: hsk1(+), a Schizosaccharomyces pombe gene related to Saccharomyces cerevisiae  
A:Reference number: S56143; MUID:95347336; PMID:7621824  
A:Accession: S56143  
A:Molecule type: DNA  
A:Residues: 1-507 <MAS>  
A:Cross-references: GB:D50493; NID:g807845; PIDN:BAA09087.1; PID:d1009722; PID:g807846  
A:Experimental source: strain JY2  
R:Lyne, M.; Rajandream, M.A.; Barrall, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.  
submitted to the EMBL Data Library, January 1999  
A:Reference number: 221944  
A:Accession: T40682  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-507 <LYN>  
A:Cross-references: EMBL:AL035263; PIDN:CAA22885.1; GSPDB:GN00067; SPDB:SPBC776.12c  
A:Experimental source: strain 972h-; cosmid c776  
C:Genetics:  
A:Gene: hsk1; SPBC776.12c  
A:Map position: 2  
A:Introns: 77/1; 359/1  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
C:Keywords: phosphotransferase; protein kinase  
F:66-404/Domain: protein kinase homology <KIN>

Query Match 7.8%; Score 83; DB 2; Length 507;  
Best Local Similarity 22.4%; Pred. No. 13;  
Matches 28; Conservative 24; Mismatches 55; Indels 18; Gaps 5;

Qy 83 LIWQTYEALYKSTNKP-----LMIHILDECPHSQALKKVFNAEKEIOKLAEOFVLL 138  
Db 380 ILWASCSGASIIYKELRHKPSQERLCDFLEKCLDCKNRISAE-----EALDHDFLYL 435

Qy 139 -NLVETTDKHLSPDQGVPRIMFVDPSTLVTRADITGRYSNRLYAY---EPADTALLLDN 194  
Db 436 DNLAYEKDDDDTAFDNS-----FGETSFEKDEDLTAKHLSHILDFKEQETDEPTSLSK 489

Qy 195 MKKAL 199  
Db 490 RKRSI 494

RESULT 14

C70441  
lipoate-protein ligase A - Aquifex aeolicus.  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: C70441  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: C70441  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-788 <AQF>  
A:Cross-references: GB:AE000748; NID:g2983960; PIDN:AAC07512.1; PID:g2983965; GB:AE00  
A:Experimental source: strain Vfs  
C:Genetics:  
A:Gene: lplA

Query Match 7.8%; Score 82.5; DB 2; Length 788;  
Best Local Similarity 23.0%; Pred. No. 25;  
Matches 49; Conservative 34; Mismatches 73; Indels 57; Gaps 10;

Qy 25 VRKSRVAMEKIPVSFLLVLSYTLARDTTVKPGAKDKDSR-PKLPTLSR-GW--G 80  
Db 585 VQKRSQAQKPVNENLKEFIPQYKVPK---VAPGQEDM--SRLPIVRETSTEIGFING 639

Qy 81 DOLIWTQYEEALYKSTNKPIMIIHILDECPHSQALKKVFNAEKEIOKLAEOFVLLNL 140  
Db 640 QKVSEFALQVAVELLSKAKRPTLVGPL-----VLWANEAKKKEGIIRLKE 689

Qy 141 VYETTDKHLSPDQGVPRIMFVDP-----PSLTV-----RADITGR 175  
Db 690 KYPNLNIHLPLD--YKPKNKKVDFSEVDPNPHLSVLGYEHDITLMIGVHCYRTDFTIR 747

Qy 176 YSNR-----LYAYEPADTALLLDNMKK 197  
Db 748 MLKKYTDTKIVTLCTLYGHDPDADVSLAGVNSKK 780

RESULT 15  
G59431  
phosphatidylinositol polyphosphate 5-phosphatase, isoform b [imported] - huam  
C:Species: Homo sapiens (man)  
C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 19-Jul-2002  
C:Accession: G59431; H59431; A59443; B59432; C59432; D59432; E59432; F59432  
R:Duoden, R.; Griffiths, G.; Frank, R.; Argos, P.; Kreis, T.E.  
Cell 64, 649-665, 1991  
A:Title: Beta-COP, a 110 kd protein associated with non-clathrin-coated vesicles and  
A:Reference number: G59431; MUID:91121521; PMID:1840503  
A:Accession: G59431  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-893 <DUD>  
A:Cross-references: GB:NP\_001578; PID:g13325070; PIDN:NP\_001578.2  
R:Mueller, O.T.; Hartsfield, J.K. Jr.; Gallardo, L.A.; Essig, Y.P.; Miller, K.L.; Pap  
Am. J. Hum. Genet. 49, 804-810, 1991  
A:Title: Lowe oculocerebrorenal syndrome in a female with a balanced X;20 translocati  
A:Reference number: H59431  
A:Accession: H59431  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-893 <MUE>  
A:Cross-references: GB:NP\_001578; PID:g13325070; PIDN:NP\_001578.2  
R:Attree, O.; Olivios, I.M.; Okabe, I.; Bailey, L.C.; Nelson, D.L.; Lewis, R.A.; McInn  
Nature 358, 239-242, 1992  
A:Title: The Lowe's oculocerebrorenal syndrome gene encodes a protein highly homologo  
A:Reference number: A59443; MUID:92334430; PMID:1321346  
A:Accession: A59443  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-893 <ATT>  
A:Cross-references: GB:NP\_001578; PID:g13325070; PIDN:NP\_001578.2

Search completed: January 17, 2003, 12:57:23  
Job time : 22 secs



GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 12:55:16 ; Search time 12 Seconds

(without alignments)  
712.010 Million cell updates/sec

Title: US-09-646-778A-288

Perfect score: 1063

Sequence: 1 RLSCAGTILSGSGPHPSRLT.....DTALLLDNMKALKLLKTEL 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	448	42.1	183	1 XAG_XENLA	P55868 xenopus lae
2	437.5	41.2	185	1 NP77_XENLA	P55869 xenopus lae
3	87.5	8.2	458	1 YMS1_YEAST	P42933 saccharomyc
4	85.5	8.0	1829	1 MYSA_CHICK	Q02440 gallus gall
5	84	7.9	811	1 RFXL_YEAST	P48743 saccharomyc
6	83.5	7.9	217	1 YD67_SCHPO	Q10319 schizosacch
7	83	7.8	507	1 HSK1_SCHPO	P50582 schizosacch
8	81.5	7.7	1398	1 OCL1_HUMAN	Q01968 homo sapien
9	81.5	7.7	1398	1 TOP2_PLAFA	P41001 plasmodium
10	80	7.5	339	1 YTH6_RHOER	P43484 rhodococcus
11	80	7.5	843	1 YMS1_YEAST	Q05050 saccharomyc
12	79.5	7.5	301	1 ERA_HELPJ	Q921w0 helicobacte
13	79	7.4	510	1 CBPL_ORISA	F37890 oryza sativ
14	78.5	7.4	280	1 AROE_BACSU	P54374 bacillus su
15	78.5	7.4	1262	1 TPP2_MOUSE	Q64514 mus musculu
16	78.5	7.4	1828	1 MYSA_RAT	Q9qyf3 rattus norv
17	78	7.3	337	1 HAIR_DROME	P14003 drosophila
18	78	7.3	887	1 ITH3_RAT	Q63416 rattus norv
19	78	7.3	1853	1 MYSA_MOUSE	Q99104 mus musculu
20	78	7.3	1855	1 MYSA_HUMAN	Q9y411 homo sapien
21	77	7.2	729	1 YP65_CAEEL	Q09214 caenorhabdi
22	76.5	7.2	1025	1 PO21_NASVI	Q03278 nasonia vit
23	76	7.1	543	1 OPPI_ECOLI	P23843 escherichia
24	76	7.1	2145	1 U520_CAEEL	Q9u2d0 caenorhabdi
25	75.5	7.1	580	1 TREZ_MYCTU	Q10769 mycobacteri
26	75.5	7.1	2318	1 NTC3_MOUSE	Q61982 mus musculu
27	75	7.1	1694	1 SN_MOUSE	Q62230 mus musculu
28	74.5	7.0	381	1 PRLP_BOVIN	Q9qkn8 bos taurus
29	74.5	7.0	905	1 ZO3_MOUSE	Q9qxl1 mus musculu
30	74.5	7.0	2319	1 NTC3_RAT	Q9r172 rattus norv
31	74.5	7.0	2321	1 NTC3_HUMAN	Q9um47 homo sapien
32	74	7.0	260	1 TATD_ECOLI	P27859 escherichia
33	74	7.0	986	1 GM13_RAT	Q62839 rattus norv

## RESULT 1

ID	XAG_XENLA	STANDARD;	PRT;	183 AA.
AC	P55868;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DE	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Putative secreted protein XAG precursor.			
GN	XAG.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cement gland.			
RX	MEDLINE=97003224; PubMed=88050563;			
RA	Sive H., Bradley L.;			
RT	"A sticky problem: the Xenopus cement gland as a paradigm for anteroposterior patterning.";			
RL	Dev. Dyn. 205:265-280 (1996).			
CC	-1- SUBCELLULAR LOCATION: Secreted (Probable).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND.			
CC	-1- SIMILARITY: HIGH; TO XENOPUS NP77.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U76752; AAB18819.1; .			
KW	Signal.			
FT	SIGNAL. 1 18 POTENTIAL.			
FT	CHAIN 19 183 PUTATIVE SECRETED PROTEIN XAG.			
SQ	SEQUENCE 183 AA; 20475 MW; 95E12B06FCB8DCE8 CRC64;			
Query Match 42.1%; Score 448; DB 1; Length 183;				
Best Local Similarity 49.4%; Pred. No. 8e-35;				
Matches 88; Conservative 29; Mismatches 51; Indels 10; Gaps 3;				
QY	39 AFLLLVALSYTLARDITVKP---GAKKDTKDSRPKLP-----QTLSRGWGQDIWTQTY 89			
DB	6 SLVCLVLLCSALGEAVLKKPKKQAGTTDTKTDQEPAPIKTKGLKTLDRNGESIEWQTY 65			
QY	90 EALYKSKTSNPLMTIIHLLDECPHQALKKVFAENKEIOKLA-EQVVLNLVYETTDKH 148			
DB	66 EEGLAKARENKPLMVIHLEDCPYSTALKAFVAPMAQKLAQEDFIMLNLPVADEN 125			
QY	149 LSPDGQYVPRIMFVDPSTLVPRADITGYRNSRLYAYEPADTALLLDNMKALKLLKTEL 206			
DB	126 QSPDGHVYPRVIFDPSLTVRSOLKGRYGNKMYAYDADDIPELITNMKAKSFLKTEL 183			

P24107 human immun  
P18948 caenorhabdi  
P39983 saccharomyc  
P34736 pichia stip  
Q9ptu3 pagrus majo  
Q8zan6 yersinia pe  
P15704 clostridium  
Q03412 caenorhabdi  
Q09246 caenorhabdi  
Q57741 methanococ  
P37512 bacillus su  
Q52675 rhodobacter

## ALIGNMENTS

```

RESULT 2
NP77_XENLA          STANDARD;          PRT;   185 AA.
ID  NP77_XENLA
AC  P55869;
DT  01-NOV-1997 (Rel. 35, Created)
DT  15-JUL-1998 (Rel. 36, Last sequence update)
DT  15-JUL-1998 (Rel. 36, Last annotation update)
DE  Putative secreted protein NP77 precursor.
GN  NP77.
OS  Xenopus laevis (African clawed frog).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC  Xenopodinae; Xenopus.
OX  NCBI_TaxID=8355;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Cement gland;
RA  Aberger F., Schuren C., Lepperdinger G., Richter K., Grunz H.;
RL  Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC  -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC  -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND
CC  (POTENTIAL).
CC  -!- SIMILARITY: HIGH, TO XENOPUS XAG.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; J082110; AAB49974.1; -
KW  Signal.
FT  SIGNAL.
FT  CHAIN.
SQ  SEQUENCE 185 AA; 20442 MW; AE3807C926044509 CRC64;

Query Match      41.2%; Score 437.5; DB 1; Length 185;
Best Local Similarity 53.5%; Pred. No. 7.7e-34;
Matches 84; Conservative 26; Mismatches 46; Indels 1; Gaps 1;

QY  51  ARDVTVPKAKKDTKDSRPKLPOTLSRGWDQLIWTQTYEEALYKSTSKNKLMIHHL 110
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  29  AARTDTNGAAKSEPAVPKVKGLTLDRGWGEDIWAQTYEEGLAKARENKPLMVIH 88
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  111 ECPHSQLKKVFAENKEIOKLA-EQFVLLNLVYETTDKHLSPDQGVYPRIMFVDP 169
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  89  DCPYSIALKKAFVADKMAQKLAQEDFIMNLVHPVADENQSPDGHVYPKGIFIDP 148
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  170 ADITGRYSNRLVAYEPADTALLDNNKKALKLLKTEL 206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  149 SDLKGKGNKLYAYDADDIPELITNNKKAKSFLKTEL 185
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
YM51_YEAST
ID  YM51_YEAST
AC  P42933;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  Hypothetical 52.2 kDa protein in SGI1-MRPL24 intergenic region.
GN  YMR191W OR YMR646.03.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [14]
RP  SEQUENCE FROM N.A.
RC  STRAIN=S288c / AB972;
RT  Pearson P., Bowman S., Barrell B.G., Rajandream M.A.;

Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 1-249 FROM N.A.
RC  STRAIN=W303;
RX  MEDLINE=95059068; PubMed=7969174;
RA  Gangloff S., McDonald J.P., Bendixen C., Arthur L., Rothstein R.;
RT  "The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA
RL  helicase homolog: a potential eukaryotic reverse gyrase.";
RN  Mol. Cell. Biol. 14:8391-8398(1994).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z47815; CAA87812.1; -
DR  EMBL; U22341; AAB60288.1; -
DR  SGD; S0004803; YMR191W.
KW  Hypothetical protein.
FT  CONFLICT 128 128      N -> Y (IN REF. 2).
FT  CONFLICT 132 135      FVTW -> LALG (IN REF. 2).
SQ  SEQUENCE 458 AA; 52183 MW; C31107D7B28DAE33 CRC64;

Query Match      8.2%; Score 87.5; DB 1; Length 458;
Best Local Similarity 20.1%; Pred. No. 1.3;
Matches 55; Conservative 35; Mismatches 74; Indels 109; Gaps 13;

QY  15  PSRLTQ-----RWKRSRVAMEKIPYSAFLLLVALS----- 47
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  82  PTRNNAVGNNWMMWRSLRVRHLRQITKSLDRLTLISLHSHGNSHOYNNRIFVTWKS 141
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  48  ----YTLARDTTVPKAKKDTK-----DSRPKLPQTLSSG-----W-----GD 81
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  142 ASTAFRRASGLTSPSLTRRGIAREDHFRPVPNVKSFASFPVPKAGPGLFTNNWTT 201
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  82  QLIWTQTYEEALYK---SKTSNKNPLMI-----IHLDCPHSQ----- 116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  202 RLLGORAYSTSIRKFTQPAVNMTISLRCFFNSLGLNQCSHNSCKAYONASNVTSK 261
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  117 -----ALKKV-----FAENKEIOKLAQFVLLNLVYETTDKHLSPDQGVY----- 156
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  262 HVQPVALKKLSQKDNFINRLELFKIMK--TONEVVDTSAYYMEKPGSVIEFTISEF 319
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  157 -----PRIMFVDPSTLVTRADITGRYSNRLIYAYE 184
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  320 NGTESAPLSFLDPSSL--ADLDEMIRN--YKYE 348
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
MY5A_CHICK
ID  MY5A_CHICK
AC  Q02440;
DT  01-JUN-1994 (Rel. 29, Created)
DT  01-JUN-1994 (Rel. 29, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin
DB  heavy chain p190) (Myosin-V).
GN  MYO5A.
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC  Gallus.
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Brain;
RX  MEDLINE=93012002; PubMed=1383040;
RA  Sanders G., Lichte B., Meyer H.E., Killmann M.W.;
RT  "cDNA encoding the chicken ortholog of the mouse dilute gene product.
    Sequence comparison reveals a myosin I subfamily with conserved C-
```



```

RESULT 6
YD67_SCHPO
ID YD67_SCHPO STANDARD; PRT; 217 AA.
AC Q10319;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C17G8.07 in chromosome I.
GN SPAC17G8.07
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford J., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,
RA Borzym K., Langer I., Beck A., Lehrach H., Wambutt R., Purnelle B.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- FUNCTION: REQUIRED FOR CHROMOSOMAL REPLICATION.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC7 SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D50493; BAA09087.1; -  
DR EMBL; AL035263; CAA22885.1; -  
DR EMBL; AL035263; CAB38417.1; -  
DR InterPro: IPR00719; Euk\_pkinase.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR ProDom; PS000001; Euk\_pkinase; 1.  
DR ProSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR ProSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR ProSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
KW Cell cycle; Cell division; Serine/threonine-protein kinase;  
KW Transferase; ATP-binding; Alternating splicing.  
FT DOMAIN 68 433 PROTEIN KINASE.  
FT NP\_BIND 74 82 ATP (BY SIMILARITY).  
FT BINDING 129 129 ATP (BY SIMILARITY).  
FT ACT\_SITE 216 216 BY SIMILARITY.  
FT VARSPLIC 92 119 MISSING (IN SHORT ISOFORM).  
SQ SEQUENCE 507 AA; 58407 MW; 04970B58218441EE CRC64;

Query Match 7.8%; Score 83; DB 1; Length 507;  
Best Local Similarity 22.4%; Pred No. 3 8;  
Matches 28; Conservative 24; Mismatches 55; Indels 18; Gaps 5;  
QY 83 LIWQTYEALYKSKTSNPK-----LMIHLDCEPHSQALKVKYFAENKETQKLAEQFVLL 138  
DB 380 ILWASCGSASIVKELRHKPSQERLCLELDCKNKRISAE-----EALDHDPLYL 435  
QY 139 -NLVYETDKHLSPDQVPRFMFVDPSTVTRADITGRYSNRLYAY---EPADTALLLDN 194  
DB 436 DNLAYEKDDDTAFDNS-----FGETSEKEDLTAKHLSHILDPEKQETDEPTSLSK 489  
QY 195 MKKAL 199  
DB 490 RKRSI 494

RESULT 8  
ID OCLRL\_HUMAN STANDARD; PRT: 901 AA.  
AC Q01568; Q15684; Q60800; Q9UJG5; Q15774; Q9UMA5;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Inositol polyphosphate 5-phosphatase OCLRL-1 (EC 3.1.3.-) (Lowe's  
DE oculocerebrorenal syndrome protein).  
GN OCLRL OR OCLRL1 OR INPP5F.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Kidney;  
RX MEDLINE=92334430; PubMed=1321346;  
RA Attree O., Olivos I.M., Okabe I., Bailey L.C., Nelson D.L.,  
RA Lewis R.A., McInnes R.R., Nussbaum R.L.;  
RT "The Lowe's oculocerebrorenal syndrome gene encodes a protein highly  
RT homologous to inositol polyphosphate-5-phosphatase.";  
RL Nature 358:239-242(1992).  
RN [2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Brain;  
RX MEDLINE=97201100; PubMed=9048911;  
RA Nussbaum R.L., Orrison B.M., Janne P.A., Charnas L.R., Chinault A.C.;  
RT "Physical mapping and genomic structure of the Lowe syndrome gene  
RT OCLRL1.";  
RL Hum. Genet. 99:145-150(1997).  
RN [3]

RP SEQUENCE OF 81-901 FROM N.A. (ISOFORMS A AND B).  
RA Pavitt R.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 426-901 FROM N.A. (ISOFORM B).  
RA Mistry S.;  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 814-843 FROM N.A.  
RX MEDLINE=93278398; PubMed=8504307;  
RA Leahy A.M., Charnas L.R., Nussbaum R.L.;  
RT "Nonsense mutations in the OCLRL-1 gene in patients with the  
RT oculocerebrorenal syndrome of Lowe.";  
RL Hum. Mol. Genet. 2:461-463(1993).  
RN [6]  
RP CHARACTERIZATION.  
RX MEDLINE=95281554; PubMed=7761412;  
RA Zhang X., Jefferson A.B., Auethavekiat V., Majerus P.W.;  
RT "The protein deficient in Lowe syndrome is a  
RT phosphatidylinositol-4,5-bisphosphate 5-phosphatase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:4853-4856(1995).  
RN [7]  
RP CHARACTERIZATION.  
RX MEDLINE=98104142; PubMed=9430698;  
RA Zhang X., Hartz P.A., Philip E., Racusen L.C., Majerus P.W.;  
RT "Cell lines from kidney proximal tubules of a patient with Lowe  
RT syndrome lack OCLRL inositol polyphosphate 5-phosphatase and  
RT accumulate phosphatidylinositol 4,5-bisphosphate.";  
RL J. Biol. Chem. 273:1574-1582(1998).  
RN [8]  
RP VARIANTS LS T-367 DEL; G-451; S-463 AND R-524.  
RX MEDLINE=97342912; PubMed=9199559;  
RA Lin T., Orrison B.M., Leahy A.M., Suchy S.F., Bernard D.J.,  
RT "Spectrum of mutations in the OCLRL gene in the Lowe oculocerebrorenal  
RT syndrome.";  
RL Am. J. Hum. Genet. 60:1384-1388(1997).  
RN [9]  
RP VARIANTS LS TYR-375; GLN-500; ASP-508 AND CYS-513.  
RX MEDLINE=98347141; PubMed=9682219;  
RA Lin T., Orrison B.M., Suchy S.F., Lewis R.A., Nussbaum R.L.;  
RT "Mutations are not uniformly distributed throughout the OCLRL gene in  
RT Lowe syndrome patients.";  
RL Mol. Genet. Metab. 64:58-61(1998).  
RN [10]  
RP VARIANTS LS GLN-500 AND GLN-524.  
RX MEDLINE=98293952; PubMed=9632163;  
RA Kawano T., Indo Y., Nakazato H., Shimadzu M., Matsuda I.;  
RT "Oculocerebrorenal syndrome of Lowe: three mutations in the OCLRL gene  
RT derived from three patients with different phenotypes.";  
RL Am. J. Med. Genet. 77:348-355(1998).  
RN [11]  
RP VARIANT LS ARG-522.  
RX MEDLINE=99002770; PubMed=9788721;  
RA Kubota T., Sakurai A., Arakawa K., Shimazu M., Wakui K., Furihata K.,  
RA Fukushima Y.;  
RT "Identification of two novel mutations in the OCLRL gene in Japanese  
RT families with Lowe syndrome.";  
RL Clin. Genet. 54:199-202(1998).  
RN [12]  
RP VARIANTS LS GLU-357; GLU-421; ASP-424 AND TYR-498.  
RX MEDLINE=20382695; PubMed=10923037;  
RA Monnier N., Sastre V., Lerouge E., Berthoin F., Lunardi J.;  
RT "OCLRL mutation analysis in French Lowe syndrome patients:  
RT implications for molecular diagnosis strategy and genetic  
RT counseling.";  
RL Hum. Mutat. 16:157-165(2000).  
RN [13]  
RP VARIANTS LS PRO-337 AND GLY-372.  
RX MEDLINE=2032168; PubMed=10767176;  
RA Roschinger W., Muntau A.C., Rudolph G., Roscher A.A.,  
RA Kammerer S.;  
RT "Carrier assessment in families with Lowe oculocerebrorenal syndrome:  
RT





Query Match		7.5%; Score 80; DB 1; Length 843;
Best Local Similarity		21.3%; Pred. No. 14;
Matches		38; Conservative 29; Mismatches 57; Indels 54; Gaps 7;
QY	20	TQGRWRKSRVAMEKIPVSFLLIVA-----LSVTLRDITVVKPGAKKDKTDS 67
DB	453	TANRHQTEKTSQEKIKAS-FDALVARMOTKVAERETLLEDTKSKETFEKKOMQOQLKDE 511
QY	68	RPKLPOTLSRGWGDLQWTQTYEALYKSKTSNKPMLIIHLDCEPHSQALKVKFAE--- 124
DB	512	KARLDQDLEE---WGKCEQDITAEKQEELKPYH-DLANNAAEAKHTLVEERD 563
QY	125	--NKEIQKLAEQFV-----LLNL--VYETTDKHLSPD 152
DB	564	EINAEISRLQDAIVDHKRKISGYGNLDLDAQNRNIREDDKLELGQTRESLESHLND 621
RESULT 12		
ERA_HELPJ		
ID	ERA_HELPJ	STANDARD; PRT; 301 AA.
AC	O92LW0;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	GTP-binding protein era homolog.	
GN	ERA OR JHP0466.	
OS	Helicobacter pylori J99 (Campylobacter pylori J99).	
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;	
OC	Helicobacter.	
OX	NCBI_TaxID=85963;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=99120557; PubMed=9923682;	
RA	Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,	
RA	Smith D.R., Noonan B., Guld B.C., deJonge B.L., Carmel G.,	
RA	Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,	
RA	Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,	
RA	Trust T.J.;	
RT	"Genomic sequence comparison of two unrelated isolates of the human	
RT	gastric pathogen Helicobacter pylori.";	
RL	Nature 387:176-180(1999).	
CC	-1- FUNCTION: BINDS BOTH GDP AND GTP, HAS AN INTRINSIC GTPASE ACTIVITY	
CC	AND IS ESSENTIAL FOR CELL GROWTH (BY SIMILARITY).	
CC	-1- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING	
CC	PROTEINS. ERA SUBFAMILY.	
CC	-1- SIMILARITY: CONTAINS 1 KH TYPE-2 DOMAIN.	
CC		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC		
DR	EMBL; AE001480; AAD06035.1; -	
DR	HSSP; P06616; 1EGA.	
DR	InterPro; IPR004044; KH_TYPE_2.	
DR	InterPro; IPR004087; KH_dom.	
DR	InterPro; IPR005225; Small_GTP.	
DR	Pfam; PF00013; KH-domain; 1.	
DR	TIGRFAMs; TIGR00231; small_GTP; 1.	
DR	TIGRFAMs; TIGR00436; era; 1.	
DR	PROSITE; PS50823; KH_TYPE_2; 1.	
KW	GTP-binding; RNA-binding; Complete proteome.	
FT	NP_BIND 12 19	
FT	NP_BIND 64 68	
FT	NP_BIND 122 125	
FT	DNA_BIND 204 280	
FT	SEQUENCE 301 AA; 34485 MW; F668C7730882EDBF CRC64;	
SQ		
Query Match,		7.5%; Score 79.5; DB 1; Length 301;
Best Local Similarity		27.8%; Pred. No. 4.2;
Matches		25; Conservative 13; Mismatches 33; Indels 19; Gaps 4;
QY	89	YFEALYKSKTSNKP-LMITHLDECPHSQALKVKFAENKIQKLAEQFVL----- 137
DB	106	YEEFL---SLCQPHILALS KIDTATHKQVLQKL---QEYQKYSQFTALVPLSAKKSQ 158
QY	138	-LNLVYETTDKHLSPDGOYVPRIMEFVDPDL 166
DB	159	NLNALLECISKHLSPSAWLFEKDLMSDEKM 188
RESULT 13		
CBP1_ORYSA		
ID	CBP1_ORYSA	STANDARD; PRT; 510 AA.
AC	P37890;	
DT	01-OCT-1994 (Rel. 30, Created)	
DT	01-OCT-1994 (Rel. 30, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Serine carboxypeptidase I precursor (EC 3.4.16.5) (Carboxypeptidase	
DE	C).	
GN	CBP1.	
OS	Oryza sativa (Rice).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
OC	Ehrhartoideae; Oryzaceae; Oryza.	
OX	NCBI_TaxID=4530;	
[1]		
RP	SEQUENCE FROM N.A.	
RX	STRAIN=cv. Yukihikari;	
RX	MEDLINE=94213891; PubMed=8161571;	
RA	Washio K., Ishikawa K.;	
RT	"Cloning and sequencing of the gene for type I carboxypeptidase in	
RT	rice.";	
RL	Biochim. Biophys. Acta 1199:311-314(1994).	
CC	-1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a	
CC	broad specificity.	
CC	-1- PTM: THREE DISULFIDE BONDS ARE PRESENT (POTENTIAL).	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.	
CC		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC		
DR	EMBL; D17586; BAA04510.1; -	
DR	PIR; S43516; S43516.	
DR	HSSP; P08819; 1WHT.	
DR	MEROPS; S10.004; -	
DR	InterPro; IPR000379; Ser_estrs_site.	
DR	InterPro; IPR001563; Serine_carbpept.	
DR	Pfam; PF00450; serine_carbpept; 1.	
DR	PRINTS; PR00724; CRBOXYPTASEC.	
DR	ProDom; PD001189; Serine_carbpept; 2.	
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.	
DR	PROSITE; PS00342; MICROBODIES_CTER; 1.	
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.	
KW	Hydrolase; Carboxypeptidase; Glycoprotein; zymogen; Signal.	
FT	SIGNAL 1 25	
FT	PROPEP 26 36	
FT	CHAIN 37 510	
FT	ACT_SITE 194 194	
FT	ACT_SITE 434 434	
FT	ACT_SITE 487 487	
FT	CARBOHYD 154 154	
FT	CARBOHYD 268 268	
FT	CARBOHYD 418 418	
FT	SITE 508 510	
SQ	SEQUENCE 510 AA; 55709 MW; 28896247FA1371CF CRC64;	



U U U

```

CC or send an email to licensee@sib-sib.ch).
CC -----
DR EMBL; X81323; CAA57103.1; -.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.090; -.
DR MGD; MGI:102724; Tpp2.
DR InterPro; IPR000209; Peptidase_s8.
DR Pfam; PF00082; Peptidase_s8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Alternative splicing.
FT ACT_SITE 44 44 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 264 264 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 449 449 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT VARSPLICE 985 997 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 1262 AA; 139878 MW; D50D22C85544B034 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 1262;
Best Local Similarity 22.9%; Pred. No. 32;
Matches 44; Conservative 31; Mismatches 78; Indels 39; Gaps 7;

QY 5 ACTLGGSGPHPSRLTQGRVRSRVAMEKIPVSAPFLLLVALSYTLARDTVPKGA--- 61
Db 970 AGSLFLSKTELKKGKAGKAGKQKGGKVPVHYL-----IPPTKIRNGSKDKE 1021
QY 62 -----KDTKDSRPKLPQTLSRGWGDLIWTQTYEALYKSKTSNKPMLI--IHLDECP 113
Db 1022 KDSEKEDLKEEFTEALRDLKIOWMTKLDSTDIYNE-LKETYPAYLPLYVARLHQLD--- 1077
QY 114 HSQALKKVFPAENKEIQKLAQFVLLNLVYETTDKHLSPDQGVYPRIMEFVDP---SLTVRA 170
Db 1078 -----AERKMKRLNEIVDAANAVI-----SHIDQALAVYIAMKTDPRDAATIKN 1124
QY 171 DITGRYSNRLYA 182
Db 1125 DMDKQKSLIDA 1136

```

Search completed: January 17, 2003, 12:57:42  
Job time : 14 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 12:55:16 ; Search time 30 seconds

(without alignments)  
1414.857 Million cell updates/sec

Title: US-09-646-778a-288

Perfect score: 1063

Sequence: 1 RLSCAGTLGGSGPHPSRLT.....DTALLDNMKKALKLLKTEL 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	899	84.6	175	4	095994
2	828	77.9	175	11	088312
3	555	52.2	165	11	08R3W7
4	555	52.2	166	4	08TD06
5	548	51.6	159	13	090Y05
6	447	42.1	185	13	042251
7	227	21.4	172	4	096H50
8	224.5	21.1	170	11	09CQU0
9	222	20.9	172	4	095881
10	199	18.7	186	5	09NA78
11	165	15.5	257	5	09N5S7
12	119	11.2	150	17	027777
13	89	8.4	950	11	092455
14	88.5	8.3	737	2	09AF04
15	87	8.2	348	16	0926X3
16	84.5	7.9	378	16	08XIU3

17	84.5	7.9	1151	16	0929J2
18	84	7.9	435	5	09BJK8
19	83.5	7.9	918	10	09CA88
20	83.5	7.9	1011	5	09U9P6
21	83.5	7.9	1409	5	09V3A7
22	83.5	7.9	1409	5	09U6I2
23	83	7.8	786	4	09H747
24	83	7.8	976	5	09U158
25	82.5	7.8	788	16	067557
26	81.5	7.7	649	10	09C5B1
27	81.5	7.7	1158	3	09UTR5
28	81.5	7.7	1721	5	08SSQ0
29	81	7.6	601	5	095998
30	81	7.6	610	5	09XVM2
31	80.5	7.6	802	4	08TB22
32	80.5	7.6	875	12	09OGG9
33	80	7.5	154	16	09PCE7
34	80	7.5	259	10	09M9H0
35	80	7.5	623	10	09FFR8
36	80	7.5	861	5	09NKK3
37	80	7.5	1051	17	08TMS3
38	79.5	7.5	277	3	013703
39	79.5	7.5	536	10	09SKT6
40	79.5	7.5	642	16	09KFD1
41	79	7.4	275	2	087445
42	79	7.4	337	5	095NH3
43	79	7.4	477	3	013745
44	79	7.4	480	5	09VMD8
45	79	7.4	710	5	08STF7

## ALIGNMENTS

RESULT 1					
095994	ID	095994	PRELIMINARY;	PRT;	175 AA.
AC	095994;				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)			
DE	Secreted CEMENT gland protein XAG-2 homolog (Anterior GRADIENT 2				
DE	(XENEPUS LAEVIS) homolog)				
GN	HAG-2/R OR HAG-2/C OR HPC8.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BREAST;				
RX	MEDLINE=99009231; PubMed=9790916;				
RA	Thompson D.A., Weigel R.J.;				
RT	"hag-2, the human homologue of the Xenopus laevis cement gland gene				
RT	XAG-2, is coexpressed with estrogen receptor in breast cancer cell				
RT	lines.";				
RL	Biochem. Biophys. Res. Commun. 251:111-116(1998).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RN	Zhang J.S., Smith D.I.;				
RA	"Human homolog of XAG is differentially expressed in tumors.";				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=COLON;				
RA	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.				
RL	[4]				
RP	SEQUENCE FROM N.A.				
RN	Zhang J.S., Smith D.I.;				
RA	"Identification of human homolog of XAG-2 over-expressed in tumors.";				
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.				
RL	EMBL; AF038451; AAC82614.1;				

10-1998

DR EMBL; AF007791; AAC77358.1; -;  
DR EMBL; AF088667; AAF22484.1; -;  
DR EMBL; BC015503; AAH54503.1; -;  
DR EMBL; AF115926; AAL54870.1; -;  
SQ SEQUENCE 175 AA; 19979 MW; F271B1BD377BEE11 CRC64;  
  
Query Match 84.6%; Score 899; DB 4; Length 175;  
Best Local Similarity 100.0%; Pred. No. 2.6e-78;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 32 MEKIPVSFAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLGRGWDQLIWTQTYEE 91  
Db 1 MEKIPVSFAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLGRGWDQLIWTQTYEE 60  
  
Qy 92 ALYKSKTSNKPMLIHHLDECPHSQALKKVFPAENKEIQKLAQFVLLNLVYETTDKHLSP 151  
Db 61 ALYKSKTSNKPMLIHHLDECPHSQALKKVFPAENKEIQKLAQFVLLNLVYETTDKHLSP 120  
  
Qy 152 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 206  
Db 121 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 175  
  
RESULT 2  
O88312 PRELIMINARY; PRT; 175 AA.  
AC O88312;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE GOB-4 protein (Anterior GRADIENT 2) (XENOPUS LAEVIS) (XENOPUS LAEVIS).  
DE LAEVIS).  
GN AGR2 OR GOB-4 OR MAG-2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BA1B/C; TISSUE=INTESTINE;  
RA Komiya T., Hirohashi S.;  
RT "Cloning of a gene, gob-4, which is expressed in intestinal goblet cells in mice."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SWISS WEBSTER;  
RX MEDLINE=99009231; PubMed=9790916;  
RA Thompson D.A., Weigel R.J.;  
RT "hAG-2, the human homologue of the xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines."  
RL Biochem. Biophys. Res. Commun. 251:111-116(1998).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher N.C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLON;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AB016592; BAA32044.1; -;  
DR EMBL; AF044262; AAC72705.1; -;  
DR EMBL; AK007677; BAB25181.1; -;  
DR EMBL; BC013334; AAH13334.1; -;  
DR MGD; MGI:1344405; Agr2.  
SQ SEQUENCE 175 AA; 19920 MW; ACC3CFE429B668CA CRC64;  
  
Query Match 77.9%; Score 828; DB 11; Length 175;  
Best Local Similarity 91.4%; Pred. No. 1.7e-71;  
Matches 160; Conservative 8; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 32 MEKIPVSFAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLGRGWDQLIWTQTYEE 91  
Db 1 MEKIPVSFAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLGRGWDQLIWTQTYEE 60  
  
Qy 92 ALYKSKTSNKPMLIHHLDECPHSQALKKVFPAENKEIQKLAQFVLLNLVYETTDKHLSP 151  
Db 61 ALYKSKTSNKPMLIHHLDECPHSQALKKVFPAENKEIQKLAQFVLLNLVYETTDKHLSP 120  
  
Qy 152 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 206  
Db 121 DGQYVPRIMFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 175  
  
RESULT 3  
Q8R3W7 PRELIMINARY; PRT; 165 AA.  
AC Q8R3W7;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Similar to anterior gradient protein 3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC023499; AAH23499.1; -;  
SQ SEQUENCE 165 AA; 19081 MW; 39PBC7DCB36E122C CRC64;  
  
Query Match 52.2%; Score 555; DB 11; Length 165;  
Best Local Similarity 64.1%; Pred. No. 2.3e-45;  
Matches 107; Conservative 28; Mismatches 20; Indels 12; Gaps 4;  
  
Qy 41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLGRGWDQLIWTQTYEEALYKSKTSN 100  
Db 10 LLLITVSSNLA--IAIK-----KEKRP--PQTLGRGWDITWTQTYEEGLFHAKSN 58  
  
Qy 101 KPLMIHHLDECPHSQALKKVFPAENKEIQKLAQFVLLNLVYETTDKHLSPDGQYVPRI 159  
Db 59 KPLMVIHLEDCQVQALKKEFAKNEIQEMAQNDFIMLNHETTDKNLSPDGQYVPRI 118  
  
Qy 160 MFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 206  
Db 119 MFVDPSTLRADITGRYSNRLYAYEPADTALLDNMKKALKRLIQSEL 165  
  
RESULT 4  
Q8TD06 PRELIMINARY; PRT; 166 AA.  
AC Q8TD06;  
ID Q8TD06



QY 66 DSRPKLPOT-----LSRGWQDLIWTQTYEALYKSKTSNKPLMII 106  
 Db 2 ETRPRLGATGLGFSFLLVSSDGHNGLGKFGDHIHW-RTLEDGKKEAASGLPLMVI 60

QY 107 HHLDECPHSQALKKVAENKEIOKLAQFVLLNLVYETTDKH--LSPDQGVVPRIMFVDP 164  
 Db 61 IHKSWGACKALKPKFAESTEISELSHNFVMVNLDEEPEKDEFSPOGGYIPRILFLDP 120

QY 165 SLTVRADITGRYSNRLYAYEPADTALLDNKKALKLL 202  
 Db 121 SGKVHPEIINENGNSPSKYFYVSAEQVVQGMKEAQLRL 158

RESULT 8  
 Q9CQ00  
 ID Q9CQ00 PRELIMINARY; PRT; 170 AA.  
 AC Q9CQ00;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 0610040B21Rik protein (RIKEN cDNA 0610040B21 gene).  
 GN 0610040B21Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO, AND KIDNEY;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK003481; BAB22811.1; -;  
 DR EMBL; AK002862; BAB22413.1; -;  
 DR EMBL; BC006857; AAH06857.1; -;  
 DR MGD; MGI:191323; 0610040B21Rik.  
 DR InterPro; IPR000886; ER.target.  
 DR InterPro; IPR000063; ThioRed.  
 DR PROSITE; PS00014; ER.TARGET; UNKNOWN\_1.  
 DR PROSITE; PS00194; THIOREDOXIN; UNKNOWN\_1.  
 SQ SEQUENCE 170 AA; 19048 MW; 5B91FC9BE12C5E44 CRC64;

Query Match 21.1%; Score 224.5; DB 11; Length 170;  
 Best Local Similarity 34.9%; Pred. No. 1,le-13;  
 Matches 58; Conservative 25; Mismatches 58; Indels 25; Gaps 4;

QY 39 AFLLLVALSYTLARDTTVPKAKKDKTSRPLKLPQTLRSRGWQDLIWTQTYEALYKSKT 98  
 Db 14 SFLLLTSS-----DGR-----TGLGKFGDHIHW-RTLEDGKKEAAA 50

QY 99 SNKPLNIHHLDECPHSQALKKVAENKEIOKLAQFVLLNL--VYETTDKHLSPDQGVV 156

Db 51 SGLPLMVIHHSWCCACKALKPKFAESTEISELSHNFVMVNLDEEPEKDEFSPOGGYI 110  
 QY 157 PRIMEVDSLTVRADITGRYSNRLYAYEPADTALLDNKKALKLL 202  
 Db 111 PRILFLDPGKVRPEIINENGNSPSKYFYVSAEQVVQGMKEAQLRL 156

RESULT 9  
 O95881  
 ID O95881 PRELIMINARY; PRT; 172 AA.  
 AC O95881;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Unknown (Hypothetical 19.2 kDa protein).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96207227; PubMed=8619474;  
 RA Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;  
 RT "A 'double adaptor' method for improved shotgun library  
 RT construction."  
 RL Anal. Biochem. 236:107-113(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97264341; PubMed=9110174;  
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
 RT "Large-scale concatenation cDNA sequencing."  
 RL Genome Res. 7:353-358(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mei G., Yu W., Gibbs R.A.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Strausberg R.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COLON;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF131758; AAD20035.1; -;  
 DR EMBL; BC001493; AAH01493.1; -;  
 DR EMBL; BC008953; AAH08953.1; -;  
 DR InterPro; IPR000063; ThioRed.  
 DR PROSITE; PS00194; THIOREDOXIN; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 172 AA; 19206 MW; 3092E9515A7C4094 CRC64;

Query Match 20.9%; Score 222; DB 4; Length 172;  
 Best Local Similarity 34.2%; Pred. No. 2e-13;  
 Matches 54; Conservative 26; Mismatches 56; Indels 22; Gaps 3;

QY 66 DSRPKLPOT-----LSRGWQDLIWTQTYEALYKSKTSNKPLMII 106  
 Db 2 ETRPRLGATGLGFSFLLVSSDGHNGLGKFGDHIHW-RTLEDGKKEAASGLPLMVI 60

QY 107 HHLDECPHSQALKKVAENKEIOKLAQFVLLNL--VYETTDKHLSPDQGVVPRIMFVDP 164  
 Db 61 IHKSWGACKALKPKFAESTEISELSHNFVMVNLDEEPEKDEFSPOGGYIPRILFLDP 120

QY 165 SLTVRADITGRYSNRLYAYEPADTALLDNKKALKLL 202  
 Db 121 SGKVHPEIINENGNSPSKYFYVSAEQVVQGMKEAQLRL 158

```

RESULT 10
Q9NA78 ID Q9NA78 PRELIMINARY; PRT; 186 AA.
AC Q9NA78;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Y57A10A.23 protein.
GN Y57A10A.23
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL117195; CAB55026.1; -.
DR InterPro; IPR000063; Thiorod.
SQ SEQUENCE 186 AA; 21467 MW; F94E13D2F773321E CRC64;

Query Match 18.7%; Score 199; DB 5; Length 186;
Best Local Similarity 30.0%; Pred. No. 3.5e-11;
Matches 51; Conservative 38; Mismatches 57; Indels 24; Gaps 6;

QY 41 LLLVALSYTLARDTTVYKPGAKKDTKDSRPKLPQTLISRGWGDLIWTQTYEALYKSKTSN 100
DB 4 LLLALL-----VSASAYAFDKIKDSIQNPLARGFGDDIAWK-WEDAIETALD 53
QY 101 KPLMIHHLDECPHSQALKVFAEN---KEIOKLAEOQVLLNLV--YETTDKHLSPDGOY 155
DB 54 KPFIHLHKSCHACKALKKTKQSNAKAFKLSHFVWVNTEDDEPPEEYRPGKY 113
QY 156 VPRIMEVDPSTVRADITGRYSNRL-----YAYEPADTALLDNMKKAL 200
DB 114 IPRLLFLDKN-----GDLLEQFNKKAENYAYYSSPADILNSMKDVLK 159

RESULT 11
Q9NS57 ID Q9NS57 PRELIMINARY; PRT; 257 AA.
AC Q9NS57;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 30.2 kDa protein.
GN F49H12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wohldmann P., Nhan M., Chissoe S., Hawkins J.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006642; AAF39830.1; -.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR000063; Thiorod.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 257 AA; 30205 MW; BF7FA62D04DBED6 CRC64;

Query Match 15.5%; Score 165; DB 5; Length 257;
Best Local Similarity 30.7%; Pred. No. 9.8e-08;
Matches 51; Conservative 35; Mismatches 58; Indels 22; Gaps 8;

QY 41 LLLVALSYTLARDTTVYKPGAKKDTKDSRPKLPQTLISRGWGDLIWTQTYEALYKSKTSN 100
DB 3 LLLLSL---LAVAVLAKESKEIKD-----LSHGYPASIEWVE-FDKAIGIAKDLN 50
QY 101 KPLMIHHLDECPHSQALK---KVFAENKEIOKLAEOQVLLNLV--YETTDKHLSPDGOY 155
DB 51 KPFIHLHKSCHACKALKKRELKSSPKTDELIILSRKPMVNVDEDEPDPKYSPOGGY 110
QY 156 VPRIMEVDPSTVRADITGRYSNRLYAYE-PADTALLDNMKKAL 199
DB 111 IPRLLFLDKNPLKNNERKYNKYPPLPAQ---IIDGMERAL 153

RESULT 12
Q27777 ID Q27777 PRELIMINARY; PRT; 150 AA.
AC Q27777;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Protein DISULPHIDE isomerase.
GN MTH1745.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=93711463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AF000930; AAB86215.1; -.
DR InterPro; IPR000063; Thiorod.
DR Pfam; PF00085; Thiorod; 1.
KW PROSITE; PS00194; THIOREDOXIN;
KW Isomerase; Complete proteome.
SQ SEQUENCE 150 AA; 16939 MW; E0672E64B7E340EB CRC64;

Query Match 11.2%; Score 119; DB 17; Length 150;
Best Local Similarity 27.1%; Pred. No. 0.0012;
Matches 35; Conservative 28; Mismatches 46; Indels 20; Gaps 4;

QY 39 AFLLLVALSYTLARDTTVYKPGAKKDTKDSRPKLPQTLISRGWGDLIWTQTYEALYKSKT 98
DB 19 AGLSLVLIIVTV-----QPRVPSLTDEKD-LKWTYEHDEAIKASR 60
QY 99 SNKPLMIHHLDECPHSQALKKVFANKEIQ-KLAEOQVLLNLVYETTDKHLSPDGOY-V 156

```

```

Db 61 TGKNVFWFVSACQKLESETLQNTVEQVRRLAEDFTIAVKIDVDTSPALSSRYRIYGV 120
QY 157 PRIMEVDPDS 165
   | : : |||
Db 121 PTVIILDPDS 129

RESULT 13
Q924S5 PRELIMINARY; PRT; 950 AA.
AC Q924S5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN LON.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Hori O., Ogawa S.;
RT "rat Lon.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064323; BAB62423.1; -.
DR MEROPS; S16.002; -.
DR InterPro; IPR003959; AAA_ATPase_centra.
DR InterPro; IPR003111; LON.
DR InterPro; IPR001984; Lon_endopep.
DR InterPro; IPR004815; Lon_fam.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02190; LON; 1.
DR TIGRFAMs; TIGR00763; Lon; 1.
DR PROSITE; PS01046; LON_SER; UNKNOWN1.
SQ SEQUENCE 950 AA; 105792 MW; D505C3D851B6F0E7 CRC64;

Query Match 8.4%; Score 89; DB 11; Length 950;
Best Local Similarity 27.5%; Pred. No. 11;
Matches 46; Conservative 27; Mismatches 72; Indels 22; Gaps 8;

QY 50 LARDTTVKPGAKKTDKSRPKLPQTLSRG---WGDOLIWTQTYE---EALYKSKTSNKPL 103
Db 204 ISROLEVEPEGLEFAENKQSRKRLKRGKEVGDELGAQPOLEMTVA--TSDFSKEVL 261

QY 104 MIHHLDECPHS--QALKKVPFAENKEIQKLAEQFVLLNLVYETTDKHLSPDQGYVPRIME 161
Db 262 MV--EVENVAHEDFQVTEEVYKALTAIEVKTIRDIILNPLVRESVLQMMQAGQV----- 314

QY 162 VDPSL---TVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTE 205
Db 315 VDNPIYLSMDGMAALTGAESHEL--QDVLEETNILKRLYKALSLLKKE 359

RESULT 14
Q9AF04 PRELIMINARY; PRT; 737 AA.
AC Q9AF04;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical 78.0 kDa protein.
OS Frankia sp. Cp11.
OG Plasmid pFQ12.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Frankineae; Frankiaceae; Frankia.
OX NCBI_TaxID=1856;
RN [1]
RP SEQUENCE FROM N.A.
RA John T.R., Rice J.M., Johnson J.D.;
RT "Analysis of pFQ12; a 22.4 kb Frankia plasmid.";
RL Can. J. Microbiol. 0:0-0(2001).
DR EMBL; AY027524; AAK20150.1; -.

```

```

DR InterPro; IPR002543; FtsK_SpoIIIE.
DR Pfam; PF01580; FtsK_SpoIIIE; 1.
KW Hypothetical protein; plasmid.
SQ SEQUENCE 737 AA; 78018 MW; DEFCEB45DFDB7C0F CRC64;

Query Match 8.3%; Score 88.5; DB 2; Length 737;
Best Local Similarity 26.3%; Pred. No. 8.5;
Matches 46; Conservative 24; Mismatches 70; Indels 35; Gaps 8;

QY 43 LVALSYTLARDTTVKPGAKKTDKSRPKLPQTLSR-----GWGDOLIWTQTYEEA 92
Db 201 LARFSEDLANDADLPAGAGVEV-----LPGTRRSVAVLEISTVDGFAEDLLFGDDYSLE 254
QY 93 LYKSKTSNKPLMIHHLDECPHSQALKKVEA---ENKEIQKLAEQFVLLNLVYETTDK-- 147
Db 255 -----TINNPLMIGRYNTQPAEAHLRETYALVYVGEQSGGKTNQIYLVLTQIARCTDVT 309

QY 148 -HLSPDQGYVPRIMEVDPSTVRADITGRYSNRLYAYEPAD---TALLDNMKK 197
Db 310 CHIDPNGGGLPR-PWMPWVEGQAT---RPAVDVAVYEPVEAERTMGALLRGMT 360

```

```

RESULT 15
Q9Z6X3 PRELIMINARY; PRT; 348 AA.
AC Q9Z6X3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PREDICTED disulfide BOND isomerase.
GN CPN0933 OR CPJ0933 OR CP0928.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AE001673; RAD19071.1; -.
DR EMBL; AE002251; AAF38711.1; -.
DR EMBL; AP002548; BAA99141.1; -.
DR PHCI-2DPAGE; Q9Z6X3; -.
DR TIGR; CP0928; -.
KW Isomerase; Complete proteome.
SQ SEQUENCE 348 AA; 40492 MW; A7E03C043D775CEC CRC64;

Query Match 8.2%; Score 87; DB 16; Length 348;

```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 17, 2003, 12:57:01 ; Search time 11 seconds  
(without alignments)  
372.270 Million cell updates/sec

Title: US-09-646-778a-288  
Perfect score: 1063  
Sequence: 1 RLSACGTLSCGPHPSRRLLT.....DTALLDNMKALKLLKTEL 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pap:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pap:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pap:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pap:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pap:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pap:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	899	84.6	175	10	US-09-792-439-2
2	555	52.2	166	9	US-10-066-500-54
3	555	52.2	166	9	US-09-956-751-1
4	555	52.2	168	9	US-10-025-380-631
5	555	52.2	168	10	US-09-922-217-631
6	555	52.2	168	10	US-09-833-263-631
7	222	20.9	172	10	US-09-954-846-1
8	81.5	7.7	901	10	US-09-892-287-5
9	81.5	7.7	968	10	US-09-969-528-7
10	78	7.3	453	9	US-09-981-353-48
11	77	7.2	1049	10	US-09-815-242-10732
12	71.5	6.7	289	10	US-09-815-242-13244
13	71.5	6.7	289	10	US-09-815-242-13582
14	71	6.7	306	9	US-10-017-910-7
15	71	6.7	529	10	US-09-849-566-6
16	71	6.7	529	10	US-09-849-566-8
17	71	6.7	529	10	US-09-907-859-6
18	71	6.7	529	10	US-09-907-859-8
19	70.5	6.6	481	10	US-09-994-185-2

20	70.5	6.6	481	12	US-10-004-139-2	Sequence 2, Appli
21	70.5	6.6	917	9	US-10-001-835-189	Sequence 189, App
22	70.5	6.6	1252	9	US-10-047-542-89	Sequence 89, Appl
23	70	6.6	185	10	US-09-881-752A-40	Sequence 40, Appl
24	70	6.6	1198	12	US-10-001-215-2	Sequence 2, Appli
25	69	6.5	1275	9	US-10-025-201-3	Sequence 3, Appli
26	69	6.5	2905	8	US-08-424-550B-401	Sequence 401, App
27	68.5	6.4	302	10	US-09-815-242-11334	Sequence 11334, A
28	68.5	6.4	456	10	US-09-861-400-12	Sequence 12, Appl
29	68.5	6.4	481	10	US-09-861-400-11	Sequence 11, Appl
30	68.5	6.4	809	9	US-10-105-695-2	Sequence 2, Appli
31	68.5	6.4	809	9	US-10-105-694-2	Sequence 2, Appli
32	68.5	6.4	809	12	US-09-747-521-2	Sequence 2, Appli
33	68.5	6.4	809	12	US-10-106-014-2	Sequence 2, Appli
34	67.5	6.3	481	10	US-09-731-872-271	Sequence 271, App
35	67	6.3	257	10	US-09-815-242-11278	Sequence 11278, A
36	67	6.3	578	9	US-09-912-672A-27	Sequence 27, Appl
37	67	6.3	3092	10	US-09-801-368-172	Sequence 172, App
38	66.5	6.3	270	10	US-09-350-874-14	Sequence 14, Appl
39	66	6.2	453	9	US-09-738-626-4378	Sequence 4378, Ap
40	66	6.2	772	9	US-09-738-626-6661	Sequence 6661, Ap
41	65.5	6.2	305	10	US-09-895-072-3	Sequence 3, Appli
42	65.5	6.2	305	10	US-09-986-552-3	Sequence 3, Appli
43	65.5	6.2	448	10	US-09-815-108-6	Sequence 6, Appli
44	65.5	6.2	472	10	US-09-815-108-5	Sequence 5, Appli
45	65.5	6.2	504	9	US-09-992-598-119	Sequence 119, App

## ALIGNMENTS

RESULT 1  
US-09-792-439-2  
; Sequence 2, Application US/09792439  
; Patent No. US20020111303A1  
; GENERAL INFORMATION:  
; APPLICANT: Boyd, Robert Simon  
; APPLICANT: Stamps, Alasdair Craig  
; APPLICANT: Terrett, Jonathan Alexander  
; APPLICANT: Tyson, Kerry Louise  
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods  
; FILE REFERENCE: 2543-1-005N  
; CURRENT APPLICATION NUMBER: US/09/792.439  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: GB 0004576.5  
; PRIOR FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: human  
US-09-792-439-2

Query Match	84.6%	Score	899	DB	10	Length	175
Best Local Similarity	100.0%	Pred. NO.	2.7e-86				
Matches	175	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
QY	32	MEKIPVSFAILLVALSYTLARDITVVKFCAKDDTKDSRKLPTQLSRGWGDLITWTYEE	91				
Db	1	MEKIPVSFAILLVALSYTLARDITVVKFCAKDDTKDSRKLPTQLSRGWGDLITWTYEE	60				
QY	92	ALYKSTSNKPLMIIHLLDCPCPSQALKKVFENKEIOKLAOFVLLNLVYETTDKHLSP	151				
Db	61	ALYKSTSNKPLMIIHLLDCPCPSQALKKVFENKEIOKLAOFVLLNLVYETTDKHLSP	120				
QY	152	DGQYVPRIMEVDFSLTVRADITGRYSNRLYAYEPADTALLDNMKALKLLKTEL	206				
Db	121	DGQYVPRIMEVDFSLTVRADITGRYSNRLYAYEPADTALLDNMKALKLLKTEL	175				

RESULT 2  
US-10-066-500-54

; Sequence 54, Application US/10066500  
; Patent No. US20020177165A1  
; GENERAL INFORMATION:  
; APPLICANT: Avi J. Ashkenazi  
; APPLICANT: Kevin P. Baker  
; APPLICANT: David A. Botstein  
; APPLICANT: Dan L. Eaton  
; APPLICANT: Luc Desnoyers  
; APPLICANT: Sherman Fong  
; APPLICANT: Wei-Qiang Gao  
; APPLICANT: Hanspeter Gerber  
; APPLICANT: Mary E. Gerritsen  
; APPLICANT: Audrey Goddard  
; APPLICANT: Paul J. Godowski  
; APPLICANT: Austin L. Gurney  
; APPLICANT: Ivar J. Kijavlin  
; APPLICANT: Jennie P. Mather  
; APPLICANT: Mary A. Napier  
; APPLICANT: James Pan  
; APPLICANT: Nicholas F. Paoni  
; APPLICANT: Margaret Ann Roy  
; APPLICANT: Timothy A. Stewart  
; APPLICANT: Daniel Tumas  
; APPLICANT: Colin K. Watanabe  
; APPLICANT: P. Mickey Williams  
; APPLICANT: William I. Wood  
; APPLICANT: Zemin Zang  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3130RIC7  
; CURRENT APPLICATION NUMBER: US/10/066,500  
; CURRENT FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 10/002,796  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059588  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/062285  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062816  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063082  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/063329  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063733  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/066364  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 60/066840  
; PRIOR FILING DATE: 1997-11-25  
; PRIOR APPLICATION NUMBER: 60/069694  
; PRIOR FILING DATE: 1997-12-16  
; PRIOR APPLICATION NUMBER: 60/074086  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/074092  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/079294  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: 60/081049  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/095998  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: 60/097000  
; PRIOR FILING DATE: 1998-08-18  
; PRIOR APPLICATION NUMBER: 60/099601  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099803  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099811  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099812  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/100858  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/101922  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/106032  
; PRIOR FILING DATE: 1998-10-28  
; PRIOR APPLICATION NUMBER: 60/109304  
; PRIOR FILING DATE: 1998-11-20  
; PRIOR APPLICATION NUMBER: 60/125778  
; PRIOR FILING DATE: 1999-03-23  
; PRIOR APPLICATION NUMBER: 60/139695  
; PRIOR FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: 60/145070  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/145698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 60/149396  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/169495  
; PRIOR FILING DATE: 1999-12-07  
; PRIOR APPLICATION NUMBER: 08/918874  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 08/933821  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 08/960507  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 09/114844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: 09/136801  
; PRIOR FILING DATE: 1998-08-19  
; PRIOR APPLICATION NUMBER: 09/136804  
; PRIOR FILING DATE: 1998-08-19  
; PRIOR APPLICATION NUMBER: 09/136828  
; PRIOR FILING DATE: 1998-08-19  
; PRIOR APPLICATION NUMBER: 09/158342  
; PRIOR FILING DATE: 1998-09-21  
; PRIOR APPLICATION NUMBER: 09/180997  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 09/202088  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: 09/254311  
; PRIOR FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/254460  
; PRIOR FILING DATE: 1999-03-09  
; PRIOR APPLICATION NUMBER: 09/254465  
; PRIOR FILING DATE: 1999-03-05  
; PRIOR APPLICATION NUMBER: 09/284663  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 09/332928  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: 09/332929  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: 09/333075  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: 09/333077  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: 09/380137  
; PRIOR FILING DATE: 1999-08-25  
; PRIOR APPLICATION NUMBER: 09/380138  
; PRIOR FILING DATE: 1999-08-25  
; PRIOR APPLICATION NUMBER: 09/380139  
; PRIOR FILING DATE: 1999-08-25  
; PRIOR APPLICATION NUMBER: 09/403296  
; PRIOR FILING DATE: 1999-10-18  
; PRIOR APPLICATION NUMBER: 09/403297  
; PRIOR FILING DATE: 1999-10-18  
; PRIOR APPLICATION NUMBER: 09/423741

; PRIOR FILING DATE: 1999-11-10  
 ; PRIOR APPLICATION NUMBER: 09/423844  
 ; PRIOR FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: 09/522342  
 ; PRIOR FILING DATE: 2000-03-09  
 ; PRIOR APPLICATION NUMBER: 09/548815  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 09/664610  
 ; PRIOR FILING DATE: 2000-09-18  
 ; PRIOR APPLICATION NUMBER: 09/665350  
 ; PRIOR FILING DATE: 2000-09-18  
 ; PRIOR APPLICATION NUMBER: 09/709238  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 09/767609  
 ; PRIOR FILING DATE: 2001-01-22  
 ; PRIOR APPLICATION NUMBER: 09/802706  
 ; PRIOR FILING DATE: 2001-03-09  
 ; PRIOR APPLICATION NUMBER: 09/808689  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 09/866028  
 ; PRIOR FILING DATE: 2001-05-25  
 ; PRIOR APPLICATION NUMBER: 09/870574  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: 09/872035  
 ; PRIOR FILING DATE: 2001-06-01  
 ; PRIOR APPLICATION NUMBER: 09/886342  
 ; PRIOR FILING DATE: 2001-06-19  
 ; PRIOR APPLICATION NUMBER: PCT/US98/14552  
 ; PRIOR FILING DATE: 1998-07-14  
 ; PRIOR APPLICATION NUMBER: PCT/US98/18824  
 ; PRIOR FILING DATE: 1998-09-10  
 ; PRIOR APPLICATION NUMBER: PCT/US98/19093  
 ; PRIOR FILING DATE: 1998-09-14  
 ; PRIOR APPLICATION NUMBER: PCT/US98/19330  
 ; PRIOR FILING DATE: 1998-09-16  
 ; PRIOR APPLICATION NUMBER: PCT/US98/19437  
 ; PRIOR FILING DATE: 1998-09-17  
 ; PRIOR APPLICATION NUMBER: PCT/US98/24855  
 ; PRIOR FILING DATE: 1998-11-20  
 ; PRIOR APPLICATION NUMBER: PCT/US98/25108  
 ; PRIOR FILING DATE: 1998-12-01  
 ; PRIOR APPLICATION NUMBER: PCT/US98/25190  
 ; PRIOR FILING DATE: 1998-11-25  
 ; PRIOR APPLICATION NUMBER: PCT/US99/05028  
 ; PRIOR FILING DATE: 1999-03-08  
 ; PRIOR APPLICATION NUMBER: PCT/US99/12252  
 ; PRIOR FILING DATE: 1999-06-02  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20111  
 ; PRIOR FILING DATE: 1999-09-01  
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594  
 ; PRIOR FILING DATE: 1999-09-08  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090  
 ; PRIOR FILING DATE: 1999-09-15  
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547

Query Match 52.2%; Score 555; DB 9; Length 166;  
 Best Local Similarity 64.7%; Pred. No. 1.6e-50;  
 Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;  
 QY 41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPOTLSRGWGDDITWVQTYEEGLFYAKSK 100  
 Db 11 LLLVTSSNLA--IAIK-----KEKRP--POTLSRGWGDDITWVQTYEEGLFYAKSK 59  
 QY 101 KPLMIHHLDCPHSOALKKVFPAENKEIOKLAE-QFVLLNLVYETTDKHLSPDQGYVPRI 159  
 Db 60 KPLMVIHLEDCQYSOALKKVFPAENKEIOKLAE-QFVLLNLVYETTDKHLSPDQGYVPRI 119  
 QY 160 MFVDPSTLVTRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 206  
 Db 120 MFVDPSTLVTRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 166

RESULT 3

US-09-956-751-1  
 ; Sequence 1, Application US/09956751  
 ; Patent No. US20020055186A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oxford Glycosciences  
 ; TITLE OF INVENTION: Detection of Peptides  
 ; FILE REFERENCE: 2754/11022US1  
 ; CURRENT APPLICATION NUMBER: US/09/956,751  
 ; CURRENT FILING DATE: 2001-09-19  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 166  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-956-751-1

Query Match 52.2%; Score 555; DB 10; Length 166;  
 Best Local Similarity 64.7%; Pred. No. 1.6e-50;  
 Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;  
 QY 41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPOTLSRGWGDDITWVQTYEEGLFYAKSK 100  
 Db 11 LLLVTSSNLA--IAIK-----KEKRP--POTLSRGWGDDITWVQTYEEGLFYAKSK 59  
 QY 101 KPLMIHHLDCPHSOALKKVFPAENKEIOKLAE-QFVLLNLVYETTDKHLSPDQGYVPRI 159  
 Db 60 KPLMVIHLEDCQYSOALKKVFPAENKEIOKLAE-QFVLLNLVYETTDKHLSPDQGYVPRI 119  
 QY 160 MFVDPSTLVTRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 206  
 Db 120 MFVDPSTLVTRADITGRYSNRLYAYEPADTALLDNMKKALKLKTTEL 166

RESULT 4

US-10-025-380-631  
 ; Sequence 631, Application US/10025380  
 ; Publication No. US20020182191A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Meagher, Madeleine Joy  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Wang, Tongtong  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: King, Gordon E.  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Clapper, Jonathan D.  
 ; APPLICANT: Skeiky, Yasir A. W.  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Vedwick Thomas S.  
 ; APPLICANT: Carter, Darick  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
 ; FILE REFERENCE: 210121.471C14  
 ; CURRENT APPLICATION NUMBER: US/10/025,380  
 ; CURRENT FILING DATE: 2001-12-19  
 ; NUMBER OF SEQ ID NOS: 1129  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 631  
 ; LENGTH: 168  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-025-380-631

Query Match 52.2%; Score 555; DB 9; Length 168;  
 Best Local Similarity 64.7%; Pred. No. 1.7e-50;  
 Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;  
 QY 41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPOTLSRGWGDDITWVQTYEEGLFYAKSK 100

```
Db 13 LLLVTSSNLA--IAIK-----KEKRP--PQTLSRGWGDDITWVQTYEGLFYAQSK 61
QY 101 KPLMIHHLEDCPHSQALKKVFQAEKQKLA--OFVLLNLVYETTDKHLSPDQGYVPRI 159
Db 62 KPLMVIHLEDCQYSQALKKVFQAEKQKLA--OFVLLNLVYETTDKHLSPDQGYVPRI 121
QY 160 MFVDPSTVTRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 206
Db 122 MFVDPSTVTRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 168

RESULT 5
US-09-922-217-631
; Sequence 631, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jlangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 631
; TYPE: PRT
; LENGTH: 168
; ORGANISM: Homo sapiens
US-09-922-217-631

Query Match 52.2%; Score 555; DB 10; Length 168;
Best Local Similarity 64.7%; Pred. No. 1.7e-50;
Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;

QY 41 LLLVALSYTLARDITVKPGAKDKTDSRPKLPQTLSRGWGDDITWVQTYEGLFYAQSK 100
Db 13 LLLVTSSNLA--IAIK-----KEKRP--PQTLSRGWGDDITWVQTYEGLFYAQSK 61
QY 101 KPLMIHHLEDCPHSQALKKVFQAEKQKLA--OFVLLNLVYETTDKHLSPDQGYVPRI 159
Db 62 KPLMVIHLEDCQYSQALKKVFQAEKQKLA--OFVLLNLVYETTDKHLSPDQGYVPRI 121
QY 160 MFVDPSTVTRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 206
Db 122 MFVDPSTVTRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 168

RESULT 6
US-09-833-263-631
; Sequence 631, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
```

```
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 631
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-833-263-631

Query Match 52.2%; Score 555; DB 10; Length 168;
Best Local Similarity 64.7%; Pred. No. 1.7e-50;
Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;

QY 41 LLLVALSYTLARDITVKPGAKDKTDSRPKLPQTLSRGWGDDITWVQTYEGLFYAQSK 100
Db 13 LLLVTSSNLA--IAIK-----KEKRP--PQTLSRGWGDDITWVQTYEGLFYAQSK 61
QY 101 KPLMIHHLEDCPHSQALKKVFQAEKQKLA--OFVLLNLVYETTDKHLSPDQGYVPRI 159
Db 62 KPLMVIHLEDCQYSQALKKVFQAEKQKLA--OFVLLNLVYETTDKHLSPDQGYVPRI 121
QY 160 MFVDPSTVTRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 206
Db 122 MFVDPSTVTRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 168

RESULT 7
US-09-954-846-1
; Sequence 1, Application US/09954846
; Patent No. US20020102654A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; Corley, Neil C.
; Guegler, Karl J.
; Patterson, Chandra
; Baughn, Mariah R.
; TITLE OF INVENTION: THIOREDOXIN PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/954,846
; FILING DATE: 17-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/107,248
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0556 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-855-0572
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT02
; CLONE: 1925679
```

Qy 55 TVKPGAKKDTKDSRPK--LPQTLRSGWGDLIWTQTYEEAL-YKS-----KTS-NKPLMII 10

Db 551 TYKDSKTDRWDSSGCKRVP-----AWCDRLWRGTNVNQLNRYRSHMLLKTSDHKPVSAL 605  
QY 107 HHY-----DECPHQAQKKVFAENKEIQKLAEE-----QFVLLNLVYETDK-- 147  
Db 606 FHIGVKVDE-----RRYKVFEDSVRIMDRMENDLPSELSRRFEFVNKFRQLQKGG 661  
QY 148 -HLSPDGQ-----YVPRIM-----FVDPSTVTRADITGRYSNRLYAYE 184  
Db 662 FQISNNGQVCFHSFIPKLNDSQYCKPWLRAPEPGLPEPNTV--DIS-----LDYVV 713  
QY 185 PADTALLDN 194  
Db 714 SKDSVTILNS 723

## RESULT 10

US-09-981-353-48  
; Sequence 48, Application US/09981353  
; Patent No. US20020160382A1  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 48  
; LENGTH: 453  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc-feature  
; OTHER INFORMATION: Incyte ID No. US20020160382A1 1376382CD1  
US-09-981-353-48

Query Match 7.3%; Score 78; DB 9; Length 453;  
Best Local Similarity 22.5%; Pred. No. 3;  
Matches 45; Conservative 34; Mismatches 79; Indels 42; Gaps 9;

QY 25 VRKSRVAMEKIPVSAPFLLLVALSYTLARDTTVKPGAKDKDSRPKLQTLRSRGWDOLI 84  
Db 187 LRKGFYSQKVTNNPRLRIISLNTLYGPNIMTNKTD-----ANQFE 231  
QY 85 WTQTYEEALYKSTNKPMLITH--LDCPHSQ-----ALKKFAE-----NKEIQKLAEE 133  
Db 232 W---LESTLNNQQNKEKVIYIIAHVPVGLPSSQNTAMREYVNEKLLIDIFQKYSVDVIG 288  
QY 134 QF---VLLNLVYETDKHLSPDGQVPRIMFVDPSTVTRADITGRYSN---RLYAYEPA 186  
Db 289 QFYGHTHRDSIMVLSDKGSPVNS-----LFVAPAVTPVKSVLEKOTNNPGRILFQYDPR 343  
QY 187 DTALLDNMKKALKLLKTEL 206  
Db 344 DYK-LDDMLQYLYNLTEANL 362

## RESULT 11

US-09-815-242-10732  
; Sequence 10732, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10732  
; LENGTH: 1049  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10732

Query Match 7.2%; Score 77; DB 10; Length 1049;  
Best Local Similarity 25.2%; Pred. No. 12;  
Matches 35; Conservative 20; Mismatches 48; Indels 36; Gaps 5;

QY 73 QTLRSRGWDQLIWTQTYEALYKSTNKNPLMIHHLDECPHQAQKKVFAENKEIQKLA 132  
Db 822 QELNRAY-ESLVTYDDYNDIVQSSTL-----NNQIQLE 855  
QY 133 EQFVLLNLVYET-----DKHLSPDGQVPRIMFVDPSTVTRADITGRYSNRLYAYEPA 187  
Db 856 BOIG---VYETVKGSLJEEPEKEGEDFTIEFYSENWTKLYDIDSTVTDQLLGTVAAN 911  
QY 188 TALLDNMKKAL-KLLKTE 205  
Db 912 SSDARDEIEKALAKLNKTE 930

## RESULT 12

US-09-815-242-13244  
; Sequence 13244, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22



Fri Jan 17 13:01:55 2003

```

; PRIOR APPLICATION NUMBER: 60/269,308
;
; PRIORITY FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13244
; LENGTH: 289
; TYPE: prt
; ORGANISM: Streptococcus pneumoniae
; US-09-815-242-13244

```

Query Match	6.7%;	Score 71.5;	DB 10;	Length 289;
Best Local Similarity	24.8%;	Pred. No. 7.7;		
Matches 30;	Conservative 20;	Mismatches 50;	Indels 21;	Gaps 3;

Qy	58	PGAKQTKD	SRPKLPQ	TLSRGQD	LIWTQT	VEEALYK	SKTSNK	PKMLII	HHLD	CEPHSA	117
		!	! ! ! ! !	!	!	!	!	!	!	!	!
Db	160	PNYELR	TRDSR	SVLPK	LS-----	YKEAVA	ASSIAN	VAAVAA	LLAGD	MVMTAG	207
		!	!	!	!	!	!	!	!	!	!
Qy	118	LKKVFA	NKSIQ	KLAEQ	FVLL-----	NLVYET	TDKLS	PDGQY	VPRI	MFVDS	171
		!	!	!	!	!	!	!	!	!	!
Db	208	IEGLP	HERYQ	DLVRE	FAMIKQ	VTKENG	AYAT---	YLSG	AGTVM	VLA	264
		!	!	!	!	!	!	!	!	!	!
Qy	172	I	172								
Db	265	L	265								

```

RESULT 13
US-09-815-242-13582
; Sequence 13582, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13582
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13582

```

Query Match	6.7%	Score 71.5;	DB 10;	Length 289;
Best Local Similarity	24.8%;	Pred. No. 7.7;		
Matches 30;	Conservative	20;	Mismatches	50;
			Indels	21;
			Gaps	3;

Qy 58 P G A K K D T K D S R P K L P Q T L S R G W G D L I W T Q T Y E E A L Y K S K T S N K P L M I I H L D E C P H S Q A 117

```

Db      160  PNYELRTRDRSRVLPKKLS-----YKEAVAASSIANVAVALLAGDMVTAGQA 200
QY      118  LKKVFAENKEIQKLAQFVLL-----NLVYETDKHLSPDGQYVPRIMFVPSLIVRAD 171
Db      208  IEGDLFHERYRQDLVREFAMIKQVTKENGAYAT---YLSGAGPTVMVLASHDKMPTIKAE 264
QY      172  I 172
Db      265  L 265

RESULT 14
US-10-017-910-7
; Sequence 7, Application US/10017910
; Patent No. US20020159570A1
; GENERAL INFORMATION:
;   APPLICANT: Choi, Yongwon
;             Wong, Brian
;             Josien, Regis
;             Steinman, Ralph
; TITLE OF INVENTION: A PROTEIN BELONGING TO THE TNF SUPERFAMILY
; INVOLVED IN SIGNAL TRANSDUCTION, NUCLEIC ACID
; METHODS OF USE THEREOF

```

NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/017,910  
FILING DATE: 14-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/447,035  
FILING DATE: 1999-11-22  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-200  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521

```

, TELE: 10332.
, INFORMATION FOR SEQ ID NO: 7:
,   SEQUENCE CHARACTERISTICS:
,     LENGTH: 306 amino acids
,     TYPE: amino acid
,     STRANDEDNESS: single
,     TOPOLOGY: linear
,     MOLECULE TYPE: protein
,     HYPOTHETICAL: NO
,     ORIGINAL SOURCE:
,       ORGANISM: Mus musculus
,       SEQUENCE DESCRIPTION: SEQ ID NO: 7:
, US-10-017-910-7

```

Query Match	6.7%	Score 71:	DB 9:	Length 306;	
Best Local Similarity	28.6%;	Pred. No. 9.4;			
Matches 30:	Conservative 13;	Mismatches 36;	Indels 26;	Gaps	6:
QY	14	HPSR-----RLTQGRWVRKSRVAMEKIPVS	AFLLLV	ALSTLARDTTVVKP	CAKDKTK 65
Db	93	HQRSNAGRNLASTSGQPVQAQSR	EA-----	SAWMTIL----	SPAADSTPDG
					VQVQQLPK 142
QY	66	-----DSRPKLPQT	LSRGW--	GDQLIWTQTYEALY	KSKTSNKP 102

Db 143 GEPETDLNPELPAHGLIGAMSGQLSWEASQEEAFRLRSGAQFSP 187

## RESULT 15

US-09-849-566-6  
 ; Sequence 6, Application US/09849566  
 ; Patent No. US20020104109A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bremel, Robert D.  
 ; Chan, Anthony W.S.  
 ; Burns, Jane C.  
 ; TITLE OF INVENTION: Methods For Creating Transgenic Animals  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medlen & Carroll, LLP  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/849,566  
 ; FILING DATE: 04-May-2001  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/821,984  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ingolia, Diane E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: WARF-02184  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 529 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 US-09-849-566-6

Query Match 6.7%; Score 71; DB 10; Length 529;  
 Best Local Similarity 23.0%; Pred.No. 20;  
 Matches 35; Conservative 18; Mismatches 47; Indels 52; Gaps 7;  
 QY 19 LTQGRVRSRVAMEKIPVSFAFLLV-----ALSYTLARDTT-----VKPG 59  
 Db 303 LCQNTWDKVS-----KEPLSLDLSVLPAPCKGMAYTVINGTLHSAHAKYIRTWIDYG 358  
 QY 60 AKKDKDSRPKLPQTLRSRGDQIWTQYE-----EALYKSKTSNKLPMII----- 106  
 Db 359 EMKEIKGGGRGEYSK-----APELLWSQWFDGPFKIGPNGLLHTGKTFKFPFLYLIGAGI 412  
 QY 107 -----HHLDEC-----PHSQALKKKVFAENKEI 128  
 Db 413 IDEDLHELDEAAPIDHQPMPDAKSVLPEDDEI 444

Search completed: January 17, 2003, 13:00:53  
 Job time : 12 secs



Db 1 MEKIPVSALLVSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLIWTQYEE 60  
QY 92 ALYKSTSNKPLMIHHLEDCPHSQALKKVFVAENKEIQKLAQFVLLNLVYETTDKHLSP 151  
Db 61 ALYKSTSNKPLMIHHLEDCPHSQALKKVFVAENKEIQKLAQFVLLNLVYETTDKHLSP 120  
QY 152 DGQYVPRIMEVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 206  
Db 121 DGQYVPRIMEVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 175

## RESULT 2

US-09-247-155-106  
; Sequence 106, Application US/09247155A  
; Patent No. 6312922  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueret, Lydie  
; TITLE OF INVENTION: Complementary DNAs  
; FILE REFERENCE: GENSET.021A  
; CURRENT APPLICATION NUMBER: US/09/247,155A  
; EARLIER FILING DATE: 1999-02-09  
; EARLIER APPLICATION NUMBER: 60/074,121  
; EARLIER FILING DATE: 1998-02-09  
; EARLIER APPLICATION NUMBER: 60/081,563  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/096,116  
; EARLIER FILING DATE: 1998-08-10  
; EARLIER APPLICATION NUMBER: 60/099,273  
; EARLIER FILING DATE: 1998-10-04  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: Patent.pm  
; SEQ ID NO 106  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -20..-1  
US-09-247-155-106

Query Match 84.6%; Score 899; DB 4; Length 175;  
Best Local Similarity 100.0%; Pred. No. 1.2e-95;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSALLVSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLIWTQYEE 91  
Db 1 MEKIPVSALLVSYTLARDTTVPKAKKDTKDSRPKLPQTLSRGWGDLIWTQYEE 60  
QY 92 ALYKSTSNKPLMIHHLEDCPHSQALKKVFVAENKEIQKLAQFVLLNLVYETTDKHLSP 151  
Db 61 ALYKSTSNKPLMIHHLEDCPHSQALKKVFVAENKEIQKLAQFVLLNLVYETTDKHLSP 120  
QY 152 DGQYVPRIMEVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 206  
Db 121 DGQYVPRIMEVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 175

## RESULT 3

US-08-916-576B-8  
; Sequence 8, Application US/08916576B  
; Patent No. 6171816  
; GENERAL INFORMATION:  
; APPLICANT: YU, GUO-LIANG  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: EBNER, REINHARD  
; APPLICANT: ENDRESS, GREGORY A.  
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

; STREET: 1100 NEW YORK AVENUE, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/916,576B  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/024,347  
; FILING DATE: 23-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0500001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 170 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-916-576B-8

Query Match 57.3%; Score 609.5; DB 4; Length 170;  
Best Local Similarity 71.3%; Pred. No. 2.5e-62;  
Matches 122; Conservative 14; Mismatches 26; Indels 9; Gaps 2;

QY 37 VSAFLLVALLVSYTLARDTTVPKAKKDTKDSRP-KLPQTLSRGWGDLIWTQYEEALYK 95  
Db 7 LGACLLLVALLVSYTLARDTTVPKAKKDTKDSRP-KLPQTLSRGWGDLIWTQYEEALYK 58  
QY 96 SKTSNKPMLMIHHLEDCPHSQALKKVFVAENKEIQKLAQFVLLNLVYETTDKHLSPDQY 155  
Db 59 AXSNKPLMIHHLEDCPYSQLKVFVAENKEIQKLAQFVLLNLVYETTDKHLSPDQY 118  
QY 156 VPRIMEVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 206  
Db 119 VPRIMEVDPSTVRADITGRYSNRLYAYEPADTALLDNMKKALKLLKTEL 169

## RESULT 4

US-09-247-155-174  
; Sequence 174, Application US/09247155A  
; Patent No. 6312922  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueret, Lydie  
; TITLE OF INVENTION: Complementary DNAs  
; FILE REFERENCE: GENSET.021A  
; CURRENT APPLICATION NUMBER: US/09/247,155A  
; CURRENT FILING DATE: 1999-02-09  
; EARLIER APPLICATION NUMBER: 60/074,121  
; EARLIER FILING DATE: 1998-02-09  
; EARLIER APPLICATION NUMBER: 60/081,563  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/096,116  
; EARLIER FILING DATE: 1998-08-10  
; EARLIER APPLICATION NUMBER: 60/099,273  
; EARLIER FILING DATE: 1998-10-04  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: Patent.pm  
; SEQ ID NO 174



Db 66 EBLAKARENKPLMVIHLEDCPTYSIALKKAFVADRMAQKLAQEDFIMNLVHPVADEN 125  
QY 149 LSPDQYVPRIMEVDPSLTVRADITGRYSNRLYAYEPADTALLDNMKKALLKTEL 206  
Db 126 QSPDGHYVPRIFIDPSLTVSRDLKGRIGNKMYAYDADDIPELITNMKKAFLKTEL 183

RESULT 7  
US-08-916-576B-4  
; Sequence 4, Application US/08916576B  
; Patent No. 6171816  
; GENERAL INFORMATION:  
; APPLICANT: YU, GUO-LIANG  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: EBNER, REINHARD  
; APPLICANT: ENDRESS, GREGORY A.  
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/916,576B  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/024,347  
; FILING DATE: 23-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0500001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 172 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-916-576B-4

Query Match 20.9%; Score 222; DB 4; Length 172;  
Best Local Similarity 34.2%; Pred. No. 1.le-17;  
Matches 54; Conservative 26; Mismatches 56; Indels 22; Gaps 3;

QY 66 DSRPKLPQT-----LSRGWDQLIWTQTYEALYKSTNKLPMII 106  
Db 2 ETRPRIGATGLGFSLLLVISDGHNGLGKFGDHIHW-RTLEDGKKEAASGLPLMVI 60

QY 107 HHLDECPHSQALKKFAENKEIOKLABOFVLLNL--VYETDKHLSPDQGVVPRIMFVDP 164  
Db 61 IHKSWGACKALPKFAESTEISELSHNFVWVNLDEEPEKDEFSPDGGYIPRILFLDP 120

QY 165 SLTVRADITGRYSNRLYAYEPADTALLDNMKKALLKTEL 202  
Db 121 SGKVHPHEIINENGPNPKYFYVSAEQVVGCMKEAQL 158

RESULT 8  
US-08-884,681-5  
; Sequence 5, Application US/08884681

; Patent No. 5955338  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: NEW HUMAN PHOSPHATIDYLINOSITOL  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/884,681  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0334 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 901 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1420920  
US-08-884-681-5

Query Match 7.7%; Score 81.5; DB 2; Length 901;  
Best Local Similarity 23.7%; Pred. No. 1.8;  
Matches 45; Conservative 32; Mismatches 46; Indels 67; Gaps 13;

QY 55 TVKPGAKDKTSRPK--LPQTLSRGWDQLIWTQTYEAL-YKS-----KTS-NKPLMII 106  
Db 476 TYKDSKTDWRDSSGKCRVP-----AWCDRIWLWRTNVNQLNYSRSHMELKTSDHKPVSA 530

QY 107 HHL-----DECPHSQALKKFAENKEIOKLABE-----QFVLLNLVYETTK-- 147  
Db 531 FHIGVKVDE----RYYKRVFSDSVRIMDRMENDFLPSLSLSRREFVFNKFRLOKKG 586

QY 148 -HLSPDQGV-----YVPRIM-----FVDPSTLVRADITGRYSNRLYAYE 184  
Db 587 FOISNNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNEIV--DIS-----LDVYV 638

QY 185 PADTALLDN 194  
Db 639 SKDSVTILNS 648

RESULT 9  
US-09-258-643-5  
; Sequence 5, Application US/09258643  
; Patent No. 6277373  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.







```

Db 187 LRKGGYSQKVTNPNLRISLNTNLYGPNIMTLNKTDP-----ANQFE 231
Qy 85 WTQTYEALYKTSKTNKPLMIHH--LDECPHSQ--ALKKVF AE-----NKEIQKLA E 133
Db 232 W---LESTLNNSQONKEKVIHIAHPVGLPSSQNTAMREYNEKLIIDFQKYSVDVIAG 288
Qy 134 QF---VLLNLVYETTKHLSQDCQYVPRIMFVDSLVTRADITGRYSN-----RLYAYEPA 186
Db 289 QFYGHTRDSIMVLSDKGSPVNS-----LFVAPAVTPVKSVLEKQTNPNPGRIFQYDPR 343
Qy 187 DTALLDNMKKALKLLKTEL 206
Db 344 DYK-LLDMLQYILNLTEANL 362

```

## RESULT 14

```

PCT-US92-05401-4
; Sequence 4, Application PC/TUS9205401
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ibor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED
; STREET: 180 VARICK STREET
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05401
; FILING DATE: 19920626
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-PPPPPT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1160 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-05401-4

```

Query Match 7.3%; Score 78; DB 5; Length 1160;

Best Local Similarity 28.4%; Pred. No. 6.7; Mismatches 17; Conservative 17; Indels 16; Gaps 4;

```

Qy 68 RPKLPQLSRGWDQLTQTYEALYKTSKTNKPLMIHHLDECPHSQALKKVF AE 127
Db 952 RPSFP-NLTSFLGCQL---ADAEAMYQN-----VDGRVSECPHYQNRPPFSREMD 999
Qy 128 IOKLAEQFVL-----NLNVYETTKHLSQDCQYVPR 158
Db 1000 LGLSPQAQVEDSRNNLVLTSSLHLSLTGCRLPK 1034

```

## RESULT 15

```

US-08-313-185-53
; Sequence 53, Application US/08313185
; Patent No. 5851763
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate

```

```

; APPLICANT: Cole, Stewart
; APPLICANT: Young, Douglas
; APPLICANT: Zhang, Ying
; APPLICANT: Honore, Nadine
; APPLICANT: Telenti, Amalio
; APPLICANT: Bodmer, Thomas
; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
; TITLE OF INVENTION: in Mycobacterium Tuberculosis
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,185
; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356.0068-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-313-185-53

```

Query Match 7.2%; Score 77; DB 2; Length 652;

Best Local Similarity 20.7%; Pred. No. 3.7; Mismatches 49; Conservative 32; Mismatches 84; Indels 72; Gaps 11;

```

Qy 12 GPHPSRRLTQGRWRKSRVAMEKIPYSALLLVALSYTLARDTTVKPKAKKDTKDSRPKL 71
Db 211 GPNHS-PLSAAAAAIRTRMGMDDETVA---LIAGGHTLKGTHGAGPASHVGPPEAPIE 266
Qy 72 PQTLSRGWD-----QLIWTQT-----YEEALYKS-----96
Db 267 AAGL--GWASSYSGYGCADAITSGEVVVTQTPTQWNFFENLFYEWVLTSPAGAEAVDG 324
Qy 97 -----KTSNKPMLIIHHLDEC-PHSQALKKVF AEKKEIQKLAEQFVLLNLVYET 144
Db 325 APDIIPDPDPSSKKRKPMTMLVTLDRFDPYEKISRRLNDPDEFAFAFAW-----FKL 378
Qy 145 TDKHLSPD---GOYVDR--IMFVDPPLTVRADI-----TGRYSNRLYAYEPADT 188
Db 379 THRDMGPKRYIGPEVPKEDLIWQDPPQYPTEDIILKAAIAASGLVSELVSAWASAST 435

```

Search completed: January 17, 2003, 12:56:55

Job time : 17 secs

